

Raúl Martín Martín

**CONSTRUCCIÓN DE DISEÑOS
ÓPTIMOS PARA MODELOS CON
VARIABLES NO CONTRABLES**

I.S.B.N. Ediciones de la UCLM
978-84-8427-562-6



Ediciones de la Universidad
de Castilla-La Mancha

Cuenca, 2007



Departamento de Matemáticas

**CONSTRUCCIÓN DE DISEÑOS
ÓPTIMOS PARA MODELOS CON
VARIABLES NO CONTROLABLES**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

AUTOR

Raúl Martín Martín

DIRECTOR

Dr. Jesús López Fidalgo

Ciudad Real, 2006

Resumen

El objetivo de esta tesis consiste en determinar diseños óptimos cuando no todas las variables explicativas están bajo el control del experimentador. Los valores de dichas variables pueden ser conocidos o desconocidos antes de que el experimento sea realizado. Para resolver este problema se consideran distintos enfoques, dependiendo de las diferentes situaciones.

Cada uno de los capítulos está estructurado en una introducción, en donde se presentan los objetivos y las principales contribuciones, el desarrollo de los mismos y una última sección dedicada a las conclusiones.

A continuación exponemos de manera resumida los principales objetivos y contribuciones de la memoria.

En el **capítulo 1** se exponen las ideas básicas y generales del diseño óptimo de experimentos y se introduce la notación que se seguirá a lo largo de la memoria. Se definen los principales criterios de optimización, a saber, $D-$, $A-$, $G-$, $E-$, D_A- y $L - A$ optimización que serán usados a lo largo de la memoria. La segunda parte de este capítulo introduce una importante herramienta en la teoría de diseño óptimo, la derivada direccional de una función criterio. Además se proporcionan algunos teoremas de optimización basados en resultados de secciones anteriores. El capítulo concluye con la presentación de algunas técnicas numéricas que buscan el óptimo a partir de uno o varios diseños iniciales.

El **capítulo 2** aborda la teoría del Diseño Óptimo cuyas variables están fuera del control del experimentador. Comienza el capítulo

considerando diseños con dos tipos de variables: una controlable y otra no controlable conocida a priori, son los diseños marginalmente restringidos (MR). A continuación se consideran los diseños con una variable controlable y otra no controlable, desconocida antes de realizar el experimento, que serán los denominados condicionalmente restringidos (CR). Por último se estudian los diseños con los tres tipos de variables, los diseños marginal y condicionalmente restringidos (MCR). Para estos tres tipos de diseños, se proporcionan los respectivos Teoremas de Equivalencia. Basándose en los resultados anteriores, López Fidalgo y Garcet Rodríguez, [Lóp04a], proponen algoritmos para el cálculo de dichos diseños que recogemos en esta memoria. Particularizamos los resultados anteriores y los algoritmos de los diseños MR, CR y MCR respecto a algunos criterios. Además proponemos una nueva cota de la eficiencia de los diseños que se obtienen, basándonos en el resultado de Atwood, [Atw69].

El **capítulo 3** propone una clase de algoritmos multiplicativos para el cálculo de diseños óptimos de los propuestos en el capítulo 2. La adopción de estos algoritmos para el cálculo de diseños óptimos cuando no todas las variables son controlables ha sido sugerida por los trabajos de Torsney y Mandal [Tor01]. Por último obtuvimos una equivalencia entre diseños MR (MCR) D - y D_s - óptimos.

El **capítulo 4** se dedica esencialmente, al estudio de los diseños D -óptimos marginalmente restringidos mediante un método secuencial que permita dirigir los experimentos según vayan apareciendo las observaciones en la variable no controlable pero conocida. Viene motivado por el ejemplo que aparece en López-Fidalgo y Garcet-Rodríguez [Lóp04a] sobre el diseño de un experimento para evaluar la capacidad predictiva de una prueba de ejercicio preoperatorio cuyo objetivo consistía en predecir la morbilidad quirúrgica en pacientes con diagnóstico de cáncer de pulmón sometidos a una resección pulmonar electiva, a partir de las características físicas del paciente, del ejercicio físico realizado y de la existencia o no de desaturación de oxígeno en sangre durante la prueba. Bajo este marco, supusimos cómo podríamos mejorar nuestro diseño si los pacientes llegan de una manera secuencial. Se proporciona un

algoritmo para el cálculo de los diseños óptimos. En un segundo momento, se plantea la construcción de diseños con distribuciones marginales discretas y continuas clásicas supuestamente conocidas.

El **capítulo 5** contiene una primera aproximación al cálculo de diseños óptimos para modelos con observaciones correladas cuando se emplean funciones de covarianza parametrizadas. Se proponen varios ejemplos reales con modelos lineales y no lineales, resueltos algunos de ellos, mediante la adaptación de un algoritmo usado para modelos con variables no controlables conocidas antes de que el experimento sea llevado a cabo. Este algoritmo es una adaptación a los diseños MR del propuesto por Brinkulov y otros [Bri80].

Summary

Most of the literature about Optimal Experimental Designs supposes that the variables which are in the model, which generates the results, are under the control of the experimenter. In this thesis we allow variables that can not be controlled by the experimenter, and we will try to provide methods to compute designs under this framework.

Chapter 1 is devoted to provide briefly the principal ideas and the most important results in the theory of Optimal Designs. This includes several criteria (D -, A -, G -, E -, D_A - and L_A - optimality), directional derivatives (Fréchet and Gâteaux), the well known Equivalence Theorem of Kiefer and Wolfowitz and its general extension provided by Whittle.

Chapter 2 explores the theory of constrain designs, specifically those in which some of the variables in the model are not under the control of the experimenter. The algorithms proposed for computing these designs are reviewed and some of them are applied according to some criteria. We introduce a new measure for the quality of a design in comparison with the optimal design. A “new” bound lower for the efficiency is also provided.

As a result of the difficulty of the algorithms which appear in the literature, Chapter 3 gives a new kind of procedure for constructing new algorithms when the factors are not controlled. These iterations are of the form: $p_j^{(r+1)} \propto p_j^{(r)} f(d_j^{(r)})$. They were first proposed by Torsney [Tor77]. These algorithms are computed following several criteria. Computational studies show fast convergence, but near the

optimum the convergence rate slows down considerably. Although monotonicity does not guarantee fast convergence, an appropriate choice of function $f(d)$ produces almost optimal designs in a few iterations. An important result of this Chapter is the equivalence between D –optimal MR or MCR designs and the appropriate D_s –optimality.

Within the framework of the uncontrolled explanatory variables, Chapter 4 deals with the problem of constructing MR D –optimal designs for linear regression models with different types of prior marginal distributions. We discuss and illustrate two techniques to compute such designs assuming the marginal design is distributed as a continuous design. The second half of this chapter considers a modification of an example proposed by López Fidalgo and Garcet Rodríguez [Lóp04a]. The example is the following: To obtain more information in order to predict complications of surgery in the treatment of lung cancer an exercise test is used. The variables considered are the percentage of the normal value of volume of expired air for each subject (x_1), decreasing of blood oxygen concentration during exercise under 90 % during the exercise test (x_2), and the exercise time in minutes (x_3). The variable x_3 is completely under control, x_1 is not subject to control, but its values are known before the exercise is carried out, and desaturation is observed during the exercise test. The target is to predict the occurrence of any postoperative cardiorespiratory complications morbi-mortality. We consider a simplified problem taking into account only the first and the third variables. Assuming that we have calculated our design, but a new patient arrives, the marginal design changes. We want to assign an exercise time in such a way that the joint design is optimal. The sequential designs are constructed in such a way that all of information in the previous runs can be used efficiently in suggesting how the next run should be conducted.

Finally we realized that in some examples where uncontrolled variables appeared and the assumption that no correlation exists between observations is not realistic. Chapter 5 is devoted to computing designs with correlated observations where different covariance

structures are studied. Also, Brimkulov's algorithm is adapted to Marginally Restricted D -Optimality and is applied to complex situations.

This brief summary will be completed with the introductions to each chapter. Finally the last chapter reviews the main findings of the thesis and a discussion of future work.

Índice general

1. Diseño de Experimentos	1
1.1. Introducción	1
1.2. Teoría de diseños para modelos lineales	3
1.3. Matriz de Información	7
1.4. Criterios de optimización	10
1.4.1. Funciones criterio	11
1.5. Condiciones de optimización	18
1.5.1. Teorema de Equivalencia	18
1.5.2. Derivada direccional	19
1.6. Aproximación al diseño óptimo	27
1.6.1. Algoritmo de Wynn-Fedorov para D -optimización	31
1.7. Modelos no lineales	40
2. Diseños con variables no controlables	43
2.1. Introducción	43
2.2. Diseños Marginalmente Restringidos	45
2.2.1. Criterios de optimización clásicos para diseños MR	53

2.3.	Diseños Condicionalmente Restringidos	64
2.3.1.	Criterios de optimalidad clásicos para diseños CR	68
2.4.	Diseños Marginal y Condicionalmente Restringidos	76
2.4.1.	Criterios de optimalidad clásicos para diseños MCR	78
3.	Algoritmos	85
3.1.	Introducción	86
3.2.	Planteamiento del problema	86
3.3.	Construcción de diseños óptimos MR	88
3.3.1.	Construcción de diseños MR para distintas funciones criterio	90
3.3.2.	Ejemplos	93
3.4.	Construcción de diseños óptimos MCR	95
3.4.1.	Construcción de diseños óptimos MCR para distintas funciones criterio	97
3.4.2.	Ejemplos	98
3.5.	Diseños óptimos para algunos casos reales	100
3.5.1.	Procedimiento de densificación para la fabri- cación de combustible nuclear	100
3.5.2.	Una prueba física para la predicción de la morbimortalidad quirúrgica de cáncer de pul- món	102
3.6.	Conclusiones	106

4. D-Óptimos Secuenciales MR	109
4.1. Introducción	110
4.2. D-óptimos MR con distribución marginal discreta .	111
4.3. D-óptimos MR con distribución marginal continua .	112
4.3.1. Discretización de la distribución de probabilidad del diseño marginal	112
4.3.2. Modelización de $p(x_1)$ por una familia paramétrica de funciones	118
4.4. Construcción Secuencial de diseños D-óptimos MR .	119
4.4.1. Un ejemplo real	120
4.5. Conclusiones	124
5. D-óptimos MR con observaciones correladas	127
5.1. Introducción	128
5.2. Planteamiento del problema	129
5.3. Estructura de la covarianza	131
5.3.1. Función potencial modificada	133
5.3.2. D -optimización y correlación	134
5.4. Regresión lineal con observaciones correladas	135
5.5. Retención radioactiva en el cuerpo humano	141
5.6. Un procedimiento iterativo	142
5.6.1. Un ejemplo real	144
5.7. Procedimiento de densificación	146
5.8. Análisis de sensibilidad	148
5.9. Conclusiones	151

6. Conclusiones e investigación futura	153
6.1. Conclusiones	153
6.2. Líneas de investigación futuras	157
6.2.1. Modelos Lineales Generalizados	157
6.2.2. Diseños con observaciones correladas	161
A. Apéndice	171
A.1. Cálculo de las derivadas locales de algunas funciones criterio	171
A.2. Código fuente para el cálculo de diseños D —óptimos MR con observaciones correladas	174
A.3. Diseño óptimo del problema de densificación en el caso correlado	180
Bibliografía	185

Lista de Tablas

1.1. Construcción de un diseño D –óptimo para el modelo de regresión lineal a través del origen.	35
2.1. Diseños óptimos MR y sus eficiencias utilizando el algoritmo que encontramos en López Fidalgo y Garcet Rodríguez [Lóp04a].	58
2.2. Diseños D-óptimos CR y sus eficiencias.	72
3.1. Número de iteraciones necesarias para conseguir . . .	94
3.2. Diseños óptimos MCR usando un algoritmo multiplicativo	99
3.3. Número de iteraciones necesarias para alcanzar . . .	99
3.4. Frecuencias de porcentajes de aditivos U_3O_8 para los 392 casos considerados en el ejemplo	102
3.5. Diseños óptimo MCR para el procedimiento de densificación	103
3.6. Diseños óptimos MCR y eficiencias de los diseños usados en la práctica	105
3.7. Número de iteraciones necesarias para que	106

4.1.	Diseños óptimos condicionales, cotas para las eficiencias y eficiencias exactas en porcentajes para el ejemplo ($t_0 = 0$, $p_0 = 1$, $t_{l+1} = 1$, $p_{l+1} = 0$ en todos los casos)	117
4.2.	Pesos del diseño secuencial en los puntos $\{0, 1\}$ en la primera columna, y en los puntos $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$ para las segunda y tercera columnas. La última columna muestra las eficiencias del diseño secuencial respecto del correspondiente D —óptimo MR.	124
5.1.	Diseños D —Óptimos MR cuando los parámetros de interés son los de la tendencia (primera fila para cada valor de θ_2 , $j = 1$) y cuando además de los parámetros del modelo, los de la función de covarianza son también de interés (segunda fila para cada valor de θ_2 , $j = 2$) y la correspondiente información.	136
5.2.	Diseños D —óptimos MR cuando los parámetros de interés son los del modelo (primera fila para cada valor de θ_2 , $j = 1$) y cuando ambos parámetros los del modelo y los de la función de covarianza son de interés (segunda fila para cada valor de θ_2 , $j = 2$) y la información correspondiente.	137
5.3.	Diseños D —óptimos MR para la estructura de covarianza 1.) cuando los parámetros de interés son los del modelo(primera fila para cada valor de θ_2 , $j = 1$) y cuando ambos parámetros los del modelo y los de la función de covarianza son de interés (segunda fila para cada valor de θ_2 , $j = 2$) y la información correspondiente.	139

5.4. Diseños D –óptimos MR para la estructura de covarianza 2.) cuando los parámetros de interés son los del modelo (primera fila para cada valor de θ_2 , $j = 1$) y cuando ambos parámetros los del modelo y los de la función de covarianza son de interés (segunda fila para cada valor de θ_2 , $j = 2$) y la información correspondiente.	139
5.5. Diseños D –óptimos MR para la estructura de covarianza Gaussiana cuando los parámetros de interés son los del modelo (primera fila para cada valor de θ_2 , $j = 1$) y cuando ambos parámetros los del modelo y los de la función de covarianza son de interés (segunda fila para cada valor de θ_2 , $j = 2$) y la información correspondiente.	140
5.6. Diseños D –óptimos MR para la estructura de covarianza exponencial cuando los parámetros de interés son los del modelo (primera fila para cada valor de θ_2 , $j = 1$) y cuando ambos parámetros los del modelo y los de la función de covarianza son de interés (segunda fila para cada valor de θ_2 , $j = 2$) y la información correspondiente.	140
5.7. D –Óptimos MR cuando los parámetros de interés son los del modelo (primera fila para cada valor de θ_2 , $j = 1$) y cuando son de interés los parámetros del modelo y de la función de covarianza (segunda fila para cada valor de θ_2 , $j = 2$) y la información correspondiente.	143
5.8. Diseño óptimo conjunto para el modelo de viento irlandés	145
5.9. D-eficiencias de diseños D –óptimos bajo la estructura de covarianza Gaussiana con valor nominal del parámetro $\theta_2^{(G)}$, cuando la verdadera estructura de covarianza es la exponencial con valor nominal del parámetro $\theta_2^{(E)}$	148

5.10. D-eficiencias bajo la suposición de que el valor del parámetro es θ_1^* para las estructuras de covarianza exponencial (E) y Gaussiana (G) cuando ambos parámetros el de la tendencia y el de la covarianza son de interés	150
5.11. D-eficiencias para el ejemplo de la Sección 5.5 cuando solamente el parámetro de la tendencia es de interés ($j = 1$) y cuando ambos el de la tendencia y el de la covarianza son de interés ($j = 2$) cuando el verdadero valor del parámetro es θ_2^* y el que fue usado para calcular el diseño óptimo fue θ_2	150
A.1. Diseño D-óptimo MR para el ejemplo de densificación	181
A.2. Diseño D-óptimo MR para el ejemplo de densificación (cont.)	182
A.3. Diseño D-óptimo MR para el ejemplo de densificación (cont.)	183

Capítulo 1

Diseño de Experimentos

1.1. Introducción

La experimentación es una herramienta fundamental en todos los campos de investigación y desarrollo. Un experimento es diseñado para responder a una gran variedad de cuestiones de interés. En general, existen diferentes posibilidades para elegir puntos en los que tomar observaciones o regiones de observación. El problema de decidir cuántas observaciones realizar y en qué puntos define un diseño.

El resultado observado de un experimento se modeliza mediante una relación funcional con los factores que en él influyen. Esta dependencia se puede escribir como:

$$y(x) = \eta(x, \theta) + \varepsilon(x),$$

siendo y una variable aleatoria, denominada variable respuesta, que depende de una o varias variables x a las que llamaremos variables explicativas, regresoras o variables controlables. Estas, pueden ser elegidas libremente por el experimentador. La función η recibe el nombre de superficie de respuesta y es una función parcialmente conocida, es decir, está dentro de un conjunto paramétrico de funciones $\eta(x) = \eta(x, \theta)$, donde los parámetros $\theta^t = (\theta_1, \dots, \theta_m) \in \mathbb{R}^m$

son desconocidos y su especificación determina totalmente a η . Por tanto, el número de parámetros del modelo es m . Por último, el modelo presenta un error aleatorio ε que engloba tanto los errores de la realización de las pruebas como la especificación del modelo.

Para caracterizar la varianza de $y(x)$ y la dependencia entre dos observaciones $y(x)$, $y(z)$, consideramos la función de covarianza:

$$\begin{aligned} E\{[y(x) - \eta(x, \theta)][y(z) - \eta(z, \theta)]\} &= E[\varepsilon(x)\varepsilon(z)] \\ &= Cov[y(x), y(z)]. \end{aligned}$$

La cuestión que surge es: ¿En qué valores de las variables controlables deberían tomarse las observaciones y en qué proporciones para obtener una estimación óptima, en algún sentido, de los parámetros desconocidos?

En primer lugar, hemos de especificar el conjunto de puntos observables donde toman valores las variables controlables. Dicho conjunto se denomina *espacio del diseño* o *dominio experimental* y se denota por χ . En la práctica χ va a ser un subconjunto compacto de un espacio euclídeo, con frecuencia un intervalo de la recta real o producto de ellos. En adelante supondremos que dicho conjunto es compacto.

La teoría general de diseños óptimos se desarrolló en un principio para modelos de regresión lineal con observaciones incorreladas, aunque en muchos ejemplos prácticos, los modelos son no lineales y en muchos problemas es más adecuado considerar observaciones correladas.

El objetivo de este capítulo, consiste en dar una descripción general de la teoría de diseños óptimos para modelos lineales con observaciones incorreladas. En las siguientes secciones, se expondrá un resumen de conceptos y resultados fundamentales de dicha teoría.

1.2. Teoría de diseños para modelos lineales

En los diseños para modelos lineales, el modelo es lineal en los parámetros desconocidos θ , esto conduce al modelo

$$y(x) = \eta(x, \theta) + \varepsilon(x) = \theta^t f(x) + \varepsilon(x). \quad (1.1)$$

Como ya hemos indicado, el error $\varepsilon(x)$ es aleatorio, y suponemos que

$$E[\varepsilon(x)] = 0 \quad \text{y} \quad Var[\varepsilon(x)] = \sigma^2(x),$$

es decir,

$$E[y(x)] = \theta^t f(x) = \sum_{j=1}^m f_j(x) \theta_j \quad \text{y} \quad Var[y(x)] = \sigma^2(x).$$

La función de superficie respuesta $\eta(x, \theta)$ es lineal en las componentes de $f^t(x) = (f_1(x), \dots, f_m(x))$, siendo esta una función vectorial continua y conocida, y $\theta^t = (\theta_1, \dots, \theta_m)$, es un vector de parámetros desconocidos que pertenece a un conjunto Θ .

Por otro lado, podemos suponer que la varianza de la respuesta, depende de las variables controlables mediante una función conocida, $\lambda(x)$,

$$Var[y(x)] = \sigma^2(x) = \frac{\sigma^2}{\lambda(x)}.$$

Al ser $\lambda(x)$ conocida, podemos reducir el correspondiente modelo de regresión mediante las transformaciones:

$$y\sqrt{\lambda(x)} \longrightarrow y \quad , \quad f(x)\sqrt{\lambda(x)} \longrightarrow f(x)$$

a un modelo homocedástico donde $Var[y(x)] = \sigma^2$. Véase Rodríguez Torreblanca y Ortiz Rodríguez, [Rod99].

La función $\lambda(x)$ además de ser conocida, tiene que ser acotada, continua, con valores reales, positiva y definida en χ . Se denomina

función de eficiencia y refleja la estructura de heterocedasticidad del modelo.

La simplicidad de las transformaciones que son necesarias para tratar el caso general, nos permite suponer, sin pérdida de generalidad en la teoría del diseño óptimo de experimentos, que $\lambda(x) = 1$, y trabajar así de manera más cómoda.

A partir de este momento trabajaremos con modelos homocedásticos. Además, utilizaremos modelos de regresión con observaciones incorreladas.

Por tanto el modelo de regresión lineal con observaciones incorreladas, viene determinado por una terna, (χ, Θ, σ) , donde χ es un espacio métrico compacto y Θ es un espacio paramétrico. Podemos conseguir que las funciones $f_1(x), \dots, f_m(x)$ formen una base de este espacio.

Para cada valor $x \in \chi$ se obtiene una observación sobre y . Debido a este control sobre la selección de x , la cuestión que surge es ¿cuántas y en qué valores de x deberían ser tomadas observaciones sobre y para obtener la mejor inferencia sobre todos o parte de los parámetros θ ? A esta selección del número de observaciones y valores de $x \in \chi$ es a lo que se llama *diseño óptimo*.

Supongamos que de antemano el número de observaciones que podemos realizar es N . Usando notación vectorial:

$$Y_N = (y(x_1), \dots, y(x_N))^t \text{ es el vector de observaciones}$$

con

$$X_N = (x_{ij}), \quad x_{ij} = f_j(x_i), \quad i = 1, \dots, N, \quad j = 1, \dots, m$$

es decir,

$$X_N = \begin{pmatrix} f_1(x_1) & \dots & f_m(x_1) \\ \dots & \dots & \dots \\ f_1(x_N) & \dots & f_m(x_N) \end{pmatrix}.$$

De Podemos escribir el modelo (1.1) matricialmente :

$$E[Y_N] = X_N \theta,$$

y la hipótesis de que las observaciones sean incorreladas, se escribe como:

$$\Sigma_N = Cov(Y | \theta, \sigma^2, X) = \sigma^2 I_N$$

siendo I_N la matriz identidad de orden N .

Los parámetros del modelo, θ , los podemos estimar mediante mínimos cuadrados, obteniendo

$$\hat{\theta} = (X_N^t X_N)^{-1} X_N^t Y_N.$$

Este estimador coincide con el mejor estimador lineal centrado (BLUE). El estimador de mínimos cuadrados es ampliamente usado debido a su estructura, ya que es independiente de la matriz de covarianzas Σ_N . La matriz de covarianzas del estimador $\hat{\theta}$ es

$$Cov(\hat{\theta}) = (X_N^t X_N)^{-1} X_N^t \Sigma_N X_N ((X_N^t X_N)^{-1})^t = \sigma^2 (X_N^t X_N)^{-1}.$$

Definición 1.1. Se denomina *diseño discreto o exacto* de N observaciones a una sucesión de N puntos de χ , $V_N = \{x_1, \dots, x_N\}$, donde podría haber repeticiones.

El problema de diseño consiste en elegir un diseño exacto de tamaño N que minimice un funcional apropiado de $Cov(\hat{\theta})$. Desde un punto de vista matemático, en vez de considerar la matriz anterior, se considera la llamada **matriz de información** que se define

$$M(V_N) = \frac{1}{N} X_N^t X_N = \frac{1}{N} \sum_{i=1}^N f(x_i) f^t(x_i).$$

Esta matriz coincide, salvo constante, con la matriz de información de Fisher.

Dadas N observaciones, debemos decidir cuántas N_i debemos tomar en x_i , con $\sum_{i=1}^k N_i = N$, de modo que la matriz $X^t X$ sea “grande” en algún sentido. Dado que los N_i deben ser enteros, esto es un problema de programación entera que en el contexto de la teoría

de diseño se denomina **problema de diseño exacto**. Dadas estas elecciones, podemos escribir la matrix $X^t X$ del siguiente modo

$$X^t X = \sum_{i=1}^k N_i f(x_i) f^t(x_i).$$

La definición anterior, sugiere la idea de asociar al diseño una medida de probabilidad discreta, definida por $\xi_N(x_i) = \frac{N_i}{N}$, $x_i \in \chi$.

La búsqueda de diseños exactos óptimos es en general difícil, pues como dijimos se trata de problemas de programación entera, y en estas condiciones, no es sencillo definir la existencia o identificar soluciones óptimas. Parece conveniente hacer, en cierto sentido, continuo el problema.

Definición 1.2. *Cualquier medida de probabilidad, ξ , sobre χ con soporte en un conjunto finito, es un diseño aproximado o asintótico.*

Es decir, si el diseño ξ tiene peso $\xi(x_i)$ en x_i , $i = 1, \dots, k$ y el número total de observaciones que se realizan es N , entonces, se toman aproximadamente $N \cdot \xi(x_i)$ observaciones en x_i , $i = 1, \dots, k$. Además, eligiendo un N suficientemente grande, podemos aproximar un diseño de estas características a uno exacto. Las aproximaciones serán mejores cuanto mayor sea N . (Véase Imhof, López Fidalgo y Wong [Imh01].)

Todavía, podríamos pensar en una definición más general del diseño, como una medida de probabilidad cualquiera. A estos diseños se les conoce como *diseños continuos* (Atkinson [Atk92]).

Haciendo uso del Teorema de Caratheodory (ver Observación 1.2) veremos que cualquier diseño óptimo continuo lo podemos “reemplazar” por una distribución de probabilidad discreta. Por este motivo, podemos suponer que el soporte del diseño óptimo es discreto.

El diseño concentrado en los puntos x_1, \dots, x_k , con pesos respectivos p_1, \dots, p_k ($0 \leq p_i \leq 1$ para $i = 1, \dots, k$; $\sum_{i=1}^k p_i = 1$) se denotará por

$$\xi = \left\{ \begin{array}{ccc} x_1 & \dots & x_k \\ p_1 & \dots & p_k \end{array} \right\},$$

y el peso de un punto x_i , se denotará también como $\xi(x_i) = p_i$. Representaremos por ξ (o p en notación vectorial) la distribución resultante sobre χ :

$$\xi \equiv p = (\xi(x_1), \dots, \xi(x_k)) = (p_1, \dots, p_k).$$

Definición 1.3. *El soporte(o espectro) de un diseño ξ es el conjunto de puntos para los cuales el diseño tiene pesos distintos de cero:*

$$S_\xi = \{x \in \chi : \xi(x) > 0\}.$$

1.3. Matriz de Información

Ya hemos definido la matriz de información para un diseño exacto. En general, considerando el diseño x_1, \dots, x_N , con medida de probabilidad, ξ_N , definimos la matriz de información como la matriz

$$M(\xi_N) = X_N^t \Sigma_N^{-1} X_N \frac{1}{N}$$

donde $\Sigma_N = \text{diag}(\sigma^2(x_1), \dots, \sigma^2(x_N))$.

Generalizando la definición anterior para diseños aproximados y continuos, definimos la matriz de información de un diseño aproximado ξ como

$$M(\xi) = \sum_{x \in X} \sigma^{-2}(x) f(x) f^t(x) \xi(x).$$

Por último, considerando a ξ como una medida de probabilidad cualquiera en χ , la *matriz de información de un diseño continuo* es

$$M(\xi) = \int_{\chi} \sigma^{-2}(x) f(x) f^t(x) \xi(dx).$$

Observación 1.1. Suponiendo la normalidad de las observaciones, la matriz de información coincide con la matriz de información de Fisher:

$$M(\xi) = -E_{\xi} \left[\frac{\partial^2 \log l(y, \gamma, \sigma^2)}{\partial \gamma_i \partial \gamma_j} \right] = E_{\xi} \left[\frac{\partial \log l(y, \gamma, \sigma^2)}{\partial \gamma_i} \frac{\partial \log l(y, \gamma, \sigma^2)}{\partial \gamma_j} \right].$$

Esta definición se podrá aplicar a modelos no lineales, con el inconveniente de que la matriz, dependerá de los parámetros que se quieran estimar.

Notación 1.1. Sea A una matriz cualquiera de orden m . Se definen los conjuntos siguientes:

$$\mathcal{M}(A) = \{Au : u \in \mathbb{R}^m\}$$

$$\mathcal{N}(A) = \{u \in \mathbb{R}^m : Au = 0\}$$

que son los subespacios imagen y núcleo de la aplicación lineal asociada a la matriz A .

El conjunto de todos los diseños en el modelo se denotará por Ξ , mientras que el conjunto de todas las matrices de información será:

$$\mathcal{M} = \{M(\xi) : \xi \in \Xi\}$$

El conjunto \mathcal{M} tiene en general una estructura más sencilla que el conjunto Ξ . De hecho, como veremos en la siguiente proposición, \mathcal{M} es un subconjunto convexo de un espacio euclídeo, el de las matrices cuadradas de orden m y simétricas.

Teorema 1.1. (Teorema de Caratheodory.) Cada elemento M de la envolvente convexa \mathcal{M} de cualquier subconjunto S de un espacio l dimensional lo podemos expresar como combinación lineal convexa de $l + 1$ elementos de S :

$$M = \sum_{i=1}^{l+1} \alpha_i s_i, \quad \sum_{i=1}^{l+1} \alpha_i = 1, \quad \alpha_i \geq 0, \quad s_i \in S.$$

Si M es un punto de la frontera de \mathcal{M} entonces $\alpha_{l+1} = 0$.

Teorema 1.2. (Propiedades de la Matriz de Información.)

Para cualquier diseño $\xi \in \Xi$ se verifica:

1. $M(\xi)$ es simétrica y semidefinida positiva.
2. $M(\xi)$ es singular si el soporte de ξ contiene menos puntos que el número de parámetros, m .
3. El conjunto de todas las matrices de información $\mathcal{M} = \{M(\xi) : \xi \in \Xi\}$ es convexo.
4. Si $f(x)$ y $\lambda(x)$ son continuas en χ , y χ es compacto entonces \mathcal{M} es compacto.
5. Para cualquier diseño, su matriz de información siempre se puede representar por:

$$M(\xi) = \sum_{i=1}^k \lambda(x_i) f(x_i) f'(x_i) p_i$$

donde $k \leq \frac{m(m+1)}{2} + 1$, $0 \leq p_i \leq 1$ y $\sum_{i=1}^k p_i = 1$.

Observación 1.2. La propiedad (5) del Teorema anterior, nos dice que para cualquier diseño, $\tilde{\xi}$, cuyo número de puntos en su soporte supere a $\frac{m(m+1)}{2} + 1$ podemos encontrar otro diseño ξ con un número de puntos en su soporte menor o igual que $\frac{m(m+1)}{2} + 1$ y tal que $M(\xi) = M(\tilde{\xi})$. De este modo, cualquier diseño óptimo continuo puede ser reemplazado por, al menos, una distribución de probabilidad discreta con soporte finito. Esto es consecuencia del Teorema de Caratheodory aplicado al espacio de las matrices simétricas de orden m .

1.4. Criterios de optimización

Entre un conjunto de diseños, no es fácil decidir cuál es “el mejor” de ellos. Para describir los resultados de un experimento, hacemos uso de $\hat{\theta}$ y de la matriz de covarianzas de los estimadores, $Cov(\hat{\theta})$, o algunas funciones de ellas.

Sabemos que entre todos los estimadores de θ , el estimador mínimo cuadrático, $\hat{\theta}$, es el mejor en el sentido de minimizar la varianza entre todos los estimadores lineales insesgados.

Consideremos diseños exactos, se trata de elegir el diseño que minimice $Cov(\hat{\theta}) \propto M^{-1}(\xi_N)$. Necesitamos por tanto elegir un criterio, una función de la inversa de la matriz de información, que nos sirva para buscar el mejor diseño bajo su punto de vista. Su elección dependerá de los intereses que se busquen al realizar el experimento, es decir buscamos el diseño que verifica:

$$\xi_N^* = \arg \min_{\xi_N} \Phi(M^{-1}(\xi_N)),$$

siendo Φ una función que responda a las necesidades del experimentador. Nos encontramos ante un problema de optimización entera:

$$\xi_N^* = \left\{ \begin{matrix} x_i^* \\ p_i^* \end{matrix} \right\}_{i=1}^{k^*} = \arg \min_{\substack{x_i, p_i, k \\ N p_i \text{ entero}}} \Phi \left[M^{-1} \left(\left\{ \begin{matrix} x_i \\ p_i \end{matrix} \right\}_{i=1}^k \right) \right].$$

Encontrar una solución a este tipo de problemas puede resultar extremadamente difícil. Para salvar, de manera relativa, este problema, usaremos los diseños aproximados que definimos anteriormente.

Denotando, como vimos, al conjunto de diseños en el modelo por

$$\Xi = \{\xi : \xi \text{ es una medida de probabilidad en } \chi\},$$

podemos escribir el problema anterior del siguiente modo:

$$\xi^* = \arg \min_{\xi \in \Xi} \Phi[M^{-1}(\xi)].$$

A la función Φ la llamamos *criterio de optimización*. Veamos ahora algunas propiedades deseables de estas funciones criterio.

1.4.1. Funciones criterio

Definición 1.4. Diremos que una función:

$$\Phi : \mathcal{M} \longrightarrow \mathbb{R} \cup \{+\infty\}$$

acotada inferiormente es una **función criterio** si se cumple que:

$$M(\xi) \geq M(\eta) \Rightarrow \Phi[M(\xi)] \leq \Phi[M(\eta)].$$

Observación 1.3. La relación entre las matrices se refiere al orden de Loewner, es decir $A \geq B \Leftrightarrow B - A$ es semidefinida positiva.

Notación 1.2. Llamaremos $\mathcal{M}^+ = \{M \in \mathcal{M} : \det M > 0\}$. $\mathcal{L}(\mathcal{M})$ es el subespacio vectorial de las matrices simétricas de orden m generado por el conjunto \mathcal{M} . Si Φ es una función criterio definimos los conjuntos:

$$\mathcal{M}_\Phi = \{M \in \mathcal{M} : \Phi(M) < \infty\}$$

$$\Xi^* = \{\xi \in \Xi : \xi \text{ es } \Phi\text{-óptimo}\}$$

Definición 1.5. Diremos que $\xi^* \in \Xi$ es un diseño Φ -**óptimo** si verifica:

$$\Phi[M(\xi^*)] = \min_{\xi \in \Xi} \Phi[M(\xi)]$$

Definición 1.6. La función Φ es convexa si dadas cualesquiera $M_1, M_2 \in \mathcal{M}$ entonces

$$\Phi[\gamma M_1 + (1 - \gamma)M_2] \leq \gamma \Phi(M_1) + (1 - \gamma)\Phi(M_2), \quad 0 \leq \gamma \leq 1.$$

Observación 1.4. Una condición suficiente para la convexidad de la función es que,

$$\left. \frac{\partial^2}{\partial \gamma^2} \Phi[\gamma M_1 + (1 - \gamma) M_2] \right|_{\gamma=0} \geq 0$$

cuando la derivada exista.

Se exponen a continuación las definiciones y propiedades de los criterios más utilizados en la práctica.

En primer lugar, consideremos el caso en el que estamos interesados por todos los parámetros del modelo. En este caso, la matriz de información debe ser no singular y por tanto definida positiva. En este contexto, consideramos los criterios más habituales: D -optimización, A -optimización, G -optimización y E -optimización.

D -optimización (Criterio del determinante)

El criterio de D -optimización viene definido por la función criterio:

$$\Phi_D[M(\xi)] = \log |M(\xi)|^{-1} = -\log |M(\xi)|, \quad \xi \in \Xi.$$

Algunas propiedades de Φ_D son:

- El valor de Φ_D es finito si y sólo si $M(\xi)$ es no singular.
- Φ_D es continua en \mathcal{M} .
- Φ_D es convexa en \mathcal{M} y estrictamente convexa en \mathcal{M}^+ .
- En las matrices en que Φ_D es finita, también es diferenciable, y su gradiente es:

$$\nabla[-\log |M(\xi)|] = -M^{-1}(\xi).$$

Además

$$\frac{\partial \Phi_D}{\partial p_i} = -f^t(x_i) M^{-1}(\xi) f(x_i).$$

Existe una interpretación estadística interesante del diseño D — óptimo. Suponiendo normalidad en los errores del modelo lineal (1.1), la región de confianza del vector de parámetros desconocidos θ describe un elipsoide de la forma:

$$(\theta - \hat{\theta})M(\xi)(\theta - \hat{\theta}) \leq \lambda,$$

siendo $\hat{\theta}$ el estimador mínimo cuadrático de θ . Como el volumen del elipsoide es proporcional a $|M^{-1}(\xi)|^{1/2}$, este criterio consiste en minimizar el volumen de dicho elipsoide.

A—optimización (Criterio de la media de las varianzas)

El criterio de A-optimización viene definido por la función criterio:

$$\Phi_A[M(\xi)] = \text{Tr} M^{-1}(\xi), \quad \xi \in \Xi.$$

Como $\text{Cov}(\hat{\theta}) \propto M^{-1}(\xi)$, entonces

$$\sum_{j=1}^m \text{Cov}(\hat{\theta}_j) \propto \text{Tr} M^{-1}(\xi),$$

de modo que este criterio minimiza el promedio de las varianzas de los estimadores de los parámetros, pero no tiene en cuenta las covarianzas entre ellos. Este criterio fue considerado por Elfving [Elf52] y Chernoff [Che53].

Algunas propiedades de Φ_A son:

- El valor de Φ_A es finito si y sólo si $M(\xi)$ es no singular.
- Φ_A es continua en \mathcal{M} .
- Φ_A es convexa en \mathcal{M} y estrictamente convexa en \mathcal{M}^+ .

- En las matrices en que Φ_A es finita, también es diferenciable, y su gradiente es:

$$\nabla[\text{Tr}M^{-1}(\xi)] = -M^{-2}.$$

Si Φ_A es diferenciable

$$\frac{\partial \Phi_A}{\partial p_i} = -f^t(x_i)M^{-2}(\xi)f(x_i).$$

E –optimización (Criterio del autovalor más pequeño)

El criterio de E -optimización viene definido por la función criterio:

$$\Phi_E[M(\xi)] = \frac{1}{\lambda_\xi} \quad \xi \in \Xi,$$

siendo λ_ξ el mínimo autovalor de $M(\xi)$.

Algunas propiedades de Φ_E son:

- El valor de Φ_E es finito si y sólo si $M(\xi)$ es no singular.
- Φ_E es continua en \mathcal{M} .
- Φ_E es convexa en \mathcal{M} y estrictamente convexa en \mathcal{M}^+ .

Si consideramos de nuevo el elipsoide de confianza de θ , como las raíces de los autovalores son proporcionales a los ejes de dicho elipsoide, el criterio E –óptimo, minimiza el mayor de los ejes.

Observación 1.5. *Podemos considerar los criterios anteriores como casos especiales de un criterio general llamado Φ_p –optimización, y se define como:*

$$\Phi_p[M(\xi)] = \left[\frac{1}{m} \sum_{i=1}^m \left(\frac{1}{\lambda_i} \right)^p \right]^{\frac{1}{p}}, \quad 0 \leq p < \infty$$

Así, para $p \rightarrow 0$, tenemos el D -óptimo, para $p \rightarrow 1$ el A -óptimo y para $p \rightarrow \infty$ el E -óptimo, siempre que $\lim_{p \rightarrow \infty} \Phi_p(\xi)$ este definido.

Considerando el modelo lineal (1.1)

$$y(x) = \eta(x, \theta) + \varepsilon$$

con $\text{Var}(y(x)) = \sigma^2$, y usando que $\text{Var}(LZ) = L\text{Var}(Z)L^t$, podemos mostrar que la varianza de la respuesta predicha en x , es decir, $\widehat{\eta}(x, \theta) = \widehat{\theta}^t f(x)$, es:

$$\text{Var}(\widehat{y}(x)) \propto d(x, \xi) = f^t(x)M^{-1}(\xi)f(x).$$

La función $d(x, \xi)$ se denomina *función de dispersión*.

Basándonos en esto, podemos definir criterios de optimización de modo que los diseños resultantes minimicen, en algún sentido, la varianza de la respuesta predicha.

G -optimización

Este criterio consiste en minimizar el valor más grande posible de la varianza en el espacio de diseño χ . Se define por la función:

$$\Phi_G[M(\xi)] = \sup_{x \in \chi} d(x, \xi).$$

Kiefer y Wolfowitz [Kie59b] prueban la equivalencia entre este criterio y el criterio de D -optimización.

Algunas propiedades de este criterio son:

- Φ_G es continua en \mathcal{M} .
- Φ_G es convexa en \mathcal{M} y estrictamente convexa en \mathcal{M}_+ .

I -optimización

Intenta minimizar el valor esperado de la varianza de la respuesta sobre χ , y se define por:

$$\Phi_I[M(\xi)] = \int_{\chi} f^t(x)M^{-1}(\xi)f(x)d\mu ,$$

siendo μ una medida de probabilidad sobre χ .

Este criterio también es conocido como V -optimización e incluso IV -optimización.

En ocasiones, no estamos interesados en todos los parámetros, sino en algunos de ellos o combinaciones de ellos. Supongamos que estamos interesados en s combinaciones lineales de los parámetros $\theta_1, \dots, \theta_m$, es decir, en $A\theta$, siendo A una matriz $s \times m$ de rango $s \leq m$.

D_A -optimización

La función criterio viene definida por

$$\Phi_{D_A}[M(\xi)] = \log \det\{AM^{-1}(\xi)A^t\}.$$

Algunas propiedades de este criterio son:

- Φ_{D_A} es continua en \mathcal{M} .
- Φ_{D_A} es convexa en \mathcal{M} y estrictamente convexa en \mathcal{M}^+ .
- Φ_{D_A} tiene derivadas parciales para pesos positivos,

$$\frac{\partial \Phi_{D_A}}{\partial p_i} = f^t(x_i)M^{-1}(\xi)A^t[AM^{-1}(\xi)A^t]^{-1}AM^{-1}(\xi)f(x_i).$$

L_A -optimización

La función criterio viene definida por

$$\Phi_{L_A}[M(\xi)] = Tr\{AM^{-1}(\xi)A^t\}.$$

Algunas propiedades de este criterio son:

- Φ_{L_A} es continua en \mathcal{M} .
- Φ_{L_A} es convexa en \mathcal{M} .
- Φ_{L_A} tiene derivadas parciales para pesos positivos,

$$\frac{\partial \Phi_{DA}}{\partial p_i} = f^t(x_i)M^{-1}(\xi)A^tAM^{-1}(\xi)f(x_i).$$

Observamos que cuando $A = c^t$, siendo c un vector de dimensión m , este criterio es el llamado c -óptimo. Este criterio fue propuesto por Elfving [Elf52], y su interés se basa en la estimación de una función lineal de los parámetros $c^t\theta$.

Dada una matriz cualquiera, A , diremos que A^- es una ***inversa generalizada*** o ***g-inversa*** cuando $AA^-A = A$. Otra definición equivalente a ésta es la siguiente: A^- es una inversa generalizada de A si A^-u satisface la ecuación $Ax = u$ para cada u de $\mathcal{M}(A)$. La inversa generalizada existe siempre, pero en general no es única. Si A es cuadrada y regular entonces la inversa generalizada coincide con la matriz inversa.

Cuando la matriz A es simétrica, existe una g-inversa muy particular, debida a Penrose, y que denotaremos por A^+ . Esta matriz es la única que verifica lo siguiente:

$$A^+u = \begin{cases} 0 & \text{si } u \in \mathcal{N}(A) \\ w_u & \text{si } u \in \mathcal{M}(A) \end{cases}$$

donde w_u es el único vector de $\mathcal{M}(A)$ tal que $Aw_u = u$.

Se define el criterio de c -optimización por:

$$\Phi_c[M(\xi)] = c^tM^-(\xi)c.$$

Una herramienta para la construcción de c -óptimos la proporciona el Teorema de Elfving mediante un procedimiento gráfico. Aunque este método es válido para cualquier dimensión, raramente es usado para modelos con más de dos parámetros, esto es debido a la dificultad de la construcción del llamado conjunto de Elfving.

Rodríguez-Díaz [Rod00], López-Fidalgo y Rodríguez-Díaz, [Lóp04d], proporcionan un procedimiento computacional para encontrar los c -óptimos usando el método de Elfving para más de dos dimensiones.

1.5. Condiciones de optimización

En esta sección determinamos condiciones para las cuales ξ^* será óptimo. Una herramienta importante será la derivada direccional. Se enuncia el Teorema de equivalencia y su extensión debida a Whittle.

1.5.1. Teorema de Equivalencia

Kiefer y Wolfowitz [Kie59b] demostraron la equivalencia entre los criterios D - y G -óptimos, cuando el modelo es homocedástico. Es decir, utilizando ambos criterios, llegamos al mismo diseño óptimo.

Teorema 1.3. (Teorema de Equivalencia.) *Las siguientes afirmaciones son equivalentes:*

1. ξ^* maximiza $|M(\xi)|$ entre todos los diseños en χ (D -optimización)

$$\max_{\xi \in \Xi} |M(\xi)| = |M(\xi^*)|.$$

2. El diseño ξ^* minimiza $\sup_{x \in \chi} d(x, \xi)$ entre todos los diseños aproximados en χ (G -optimización)

$$\min_{\xi \in \Xi} \sup_{x \in \chi} d(x, \xi) = \sup_{x \in \chi} d(x, \xi^*).$$

Además la expresión anterior es igual a m (número de parámetros)

$$\sup_{x \in \chi} d(x, \xi^*) = m.$$

Demostración. Puede encontrarse en Kiefer y Wolfowitz [Kie59b] y en Fedorov [Fed72].

Observación 1.6. *El Teorema establece la equivalencia entre D -optimización, que se utiliza para optimizar la estimación de los parámetros del modelo, y G -optimización, donde el interés se centra en la predicción óptima de la respuesta. Esta equivalencia se da para diseños continuos **pero para diseños exactos** (para un N dado) **puede que no se verifique.***

Este Teorema, juega un papel importante en la construcción de diseños óptimos y en la realización de algoritmos para su construcción, puesto que pueden intercambiarse el uso de las propiedades de los diseños D - y G - óptimos.

El Teorema también proporciona un método simple para verificar la D -optimización de un diseño. Según el punto 2 del Teorema, basta ver que se cumple la condición de que $\sup_{x \in \chi} d(x, \xi^) = m$.*

Del Teorema se deduce que en los puntos del soporte de un diseño D - (G -óptimo) la función de dispersión $d(x, \xi^)$ alcanza su máximo valor, es decir, m . La utilidad de esta observación está en que, en lugar de comprobar que $d(x, \xi) \leq m \ \forall x \in \chi$, para saber si el diseño ξ es D -óptimo, se podría empezar comprobando si $d(x_i, \xi) = m$ en los puntos x_i del soporte de ξ . Puesto que la condición establecida es necesaria pero no suficiente, solamente serviría, en principio, para asegurar que el diseño no es D -óptimo si la condición no se verifica.*

1.5.2. Derivada direccional

Whittle [Whi73b] y Kiefer [Kie74] amplían el Teorema de equivalencia para funciones criterio más generales y a su vez White [Whi73a] lo extiende a diseños para modelos no lineales. Se introducen una serie de conceptos para enunciar el Teorema de equivalencia dado por Whittle.

Definición 1.7. Dada una función criterio $\Phi : \Xi \longrightarrow \mathbb{R} \cup \{\infty\}$ convexa, definimos la derivada direccional de Φ en $M_1 = M(\xi)$ en la dirección de $M_2 = M(\eta)$ como:

$$\begin{aligned} F_\Phi(M(\xi), M(\eta)) &= \partial\Phi(M_1, M_2) = \\ &= \lim_{\beta \rightarrow 0^+} \frac{\Phi[(1 - \beta)M_1 + \beta M_2] - \Phi(M_1)}{\beta}. \end{aligned}$$

En ocasiones escribiremos $\Phi(\xi)$ en lugar de $\Phi[M(\xi)]$ y $F_\Phi(\xi, \eta)$ en vez de $F_\Phi(M(\xi), M(\eta))$ cuando no exista confusión.

La derivada direccional puede existir incluso si Φ no es diferenciable. (Whittle [Whi73b]).

Whittle [Whi71] y Kiefer [Kie74] usan en su teoría de diseño una definición alternativa a la derivada direccional, **la derivada de Gâteaux**.

Definición 1.8. Dada una función criterio $\Phi : \Xi \longrightarrow \mathbb{R} \cup \{\infty\}$ convexa, definimos la derivada de Gâteaux de Φ en $M_1 = M(\xi)$ en la dirección de $M_2 = M(\eta)$ como:

$$G_\Phi(M(\xi), M(\eta)) = \lim_{\beta \rightarrow 0^+} \frac{\Phi(M_1 + \beta M_2) - \Phi(M_1)}{\beta}.$$

Podemos ver fácilmente la relación entre ambas definiciones:

$$F_\Phi[M(\xi), M(\eta)] = G_\Phi[M(\xi), M(\eta) - M(\xi)],$$

y

$$G_\Phi[M(\xi), M(\eta)] = F_\Phi[M(\xi), M(\xi) + M(\eta)].$$

La diferenciabilidad de Φ implica que G_Φ sea lineal en su segundo argumento, Rockafellar [Roc72].

Observamos que $G_\Phi(\xi, e_j) = \frac{\partial \Phi}{\partial p_j}$, siendo $\xi = (p_1, \dots, p_k)$ y $e_j = (0, \dots, 1, \dots, 0)$.

Lema 1.1. (*Propiedades de la derivada direccional.*)

Sea Φ convexa, entonces se verifica:

1. $F_{\Phi}(M, M) = 0, \quad \forall M \in \mathcal{M}^+$
2. $F_{\Phi}(M_1, M_2) \leq \Phi(M_2) - \Phi(M_1), \quad \forall M_1 \in \mathcal{M}^+, M_2 \in \mathcal{M}$

Demostración.

1. Esta es inmediata:

$$\begin{aligned} F_{\Phi}(M, M) &= \lim_{\beta \rightarrow 0^+} \frac{\Phi[(1 - \beta)M + \beta M] - \Phi(M)}{\beta} \\ &= \lim_{\beta \rightarrow 0^+} \frac{\Phi(M) - \Phi(M)}{\beta} = 0 \end{aligned}$$

2. Por ser Φ convexa, se verifica que:

$$\begin{aligned} \frac{\Phi[(1 - \beta)M_1 + \beta M_2] - \Phi(M_1)}{\beta} &\leq \frac{(1 - \beta)\Phi(M_1) + \beta\Phi(M_2) - \Phi(M_1)}{\beta} \\ &= \Phi(M_2) - \Phi(M_1). \end{aligned}$$

Tomando $\lim_{\beta \rightarrow 0^+}$ en la desigualdad anterior, se tiene que

$$F_{\Phi}(M_1, M_2) \leq \Phi(M_2) - \Phi(M_1)$$

□

Teorema 1.4. Sea Φ una función convexa sobre \mathcal{M} . Se verifica que

$$\xi^* \text{ es } \Phi\text{-óptimo si y sólo si } F_{\Phi}[M(\xi^*), M(\xi)] \geq 0 \quad \forall \xi \in \Xi$$

Demostración.

Si Φ se minimiza en $M(\xi^*)$ entonces esto implica que:

$$\Phi[(1 - \beta)M(\xi^*) + \beta M(\xi)] - \Phi[M(\xi^*)] \geq 0$$

para todo $\beta \in [0, 1]$ y todo $\xi \in \Xi$.

Como $(1 - \beta)M(\xi^*) + \beta M(\xi) = M[(1 - \beta)\xi^* + \beta\xi]$, por la definición de derivada direccional tenemos que:

$$F_\Phi(M_1, M_2) = \lim_{\beta \rightarrow 0^+} \frac{\Phi[(1 - \beta)M_1 + \beta M_2] - \Phi(M_1)}{\beta}.$$

Sustituyendo M_1 por $M(\xi^*)$ y M_2 por $M(\xi)$, como el numerador es positivo, se verifica que:

$$F_\Phi[M(\xi^*), M(\xi)] = \partial\Phi[M(\xi^*), M(\xi)] \geq 0 \quad \forall \xi \in \Xi,$$

y queda probada la condición necesaria.

Para ver la suficiencia, supongamos que $F_\Phi[M(\xi^*), M(\xi)] \geq 0 \quad \forall \xi \in \Xi$. Usando el Lema (1.1), obtenemos que $\Phi[M(\xi)] - \Phi[M(\xi^*)] \geq 0$ para todo $\xi \in \Xi$, es decir ξ^* es óptimo.

□

Corolario 1.1. *Bajo las hipótesis del Teorema anterior*

$$\inf_{\xi} F_\Phi[M(\xi^*), M(\xi)] = 0.$$

Demostración.

Debido a que $F_\Phi[M(\xi^*), M(\xi)] \geq 0$ para todo $\xi \in \Xi$ eligiendo $\xi = \xi^*$, concluimos, puesto que $F_\Phi[M(\xi^*), M(\xi^*)] = 0$ (Lema 1.1).

□

Teorema 1.5. (*Teorema general de equivalencia.*) Sea Φ una función criterio convexa y ξ^* un diseño tal que:

$$F_\Phi[M(\xi^*), M(\xi)] > -\infty, \quad \xi \in \Xi.$$

Entonces son equivalentes:

1. $\xi^* \in \Xi$ minimiza Φ sobre Ξ , es decir, ξ^* es Φ -óptimo.
2. ξ^* maximiza el $\inf_{\eta \in \Xi} F_\Phi(M(\xi), M(\eta))$.
3. $\inf_{\eta \in \Xi} F_\Phi(M(\xi^*), M(\eta)) = 0$.

Demostración. Puede encontrarse en Pázman [Paz86].

Aunque como se ha mencionado, no es necesario que la función sea diferenciable en un punto para que las derivadas direccionales estén bien definidas, que la función criterio Φ sea diferenciable juega un importante papel a la hora de simplificar los resultados.

Definición 1.9. La derivada direccional F_Φ se dice lineal si se verifica que $\forall \xi \in \Xi$ y $\forall \eta \in \Xi$

$$F_\Phi(\xi, \eta) = \int_{\chi} F_\Phi(\xi, \xi_x) \eta(dx),$$

donde ξ_x es el diseño unipuntual con peso 1 en $x \in \chi$.

Lema 1.2. Sea Φ convexa y sea $M \in \mathcal{M}^+$. Una condición necesaria y suficiente para que Φ sea diferenciable en M es que la derivada direccional $F_\Phi(M, \cdot)$ sea lineal.

Demostración. Una demostración de este Lema puede encontrarse en Pázman [Paz86].

Lema 1.3. *Si F_Φ es lineal se verifica que*

$$\inf_{\eta \in \Xi} F_\Phi(\xi, \eta) = \inf_{x \in \chi} F_\Phi(\xi, \xi_x).$$

Demostración.

Como $\{\xi_x : x \in \chi\} \subset \Xi$ entonces

$$\inf_{\eta \in \Xi} F_\Phi(\xi, \eta) \leq \inf_{x \in \chi} F_\Phi(\xi, \xi_x).$$

Por otro lado debido a que η es una medida de probabilidad:

$$\begin{aligned} \inf_{\eta \in \Xi} F_\Phi(\xi, \eta) &= \inf_{\eta \in \Xi} \int_{\chi} F_\Phi(\xi, \xi_x) \eta(dx) \geq \inf_{\eta \in \Xi} \int_{\chi} \inf_{x \in \chi} F_\Phi(\xi, \xi_x) \eta(dx) \\ &= \inf_{x \in \chi} F_\Phi(\xi, \xi_x). \end{aligned}$$

□

Lema 1.4. *Si F_Φ es lineal, entonces*

$$S_{\xi^*} \subseteq \{x \in \chi : F_\Phi(\xi^*, \xi_x) = 0\}.$$

Demostración.

Como ξ^* es óptimo, la derivada direccional en ξ^* para cualquier dirección es no negativa, en particular

$$\forall x \in \chi : F_\Phi(\xi^*, \xi_x) \geq 0.$$

Y por la linealidad de F_Φ se tiene que

$$0 = F_\Phi(\xi^*, \xi^*) = \int_{\chi} F_\Phi(\xi^*, \xi_x) \xi^*(dx).$$

Por tanto, como ξ^* es una medida de probabilidad la integral anterior es cero si

$$S_{\xi^*} \subseteq \{x \in \chi : F_{\Phi}(\xi^*, \xi_x) = 0\}.$$

□

Observación 1.7.

1. Si $\Phi[(1-\beta)\xi + \beta\eta]$ es diferenciable podemos escribir mediante la regla de L'Hôpital

$$F_{\Phi}(\xi, \eta) = \lim_{\beta \rightarrow 0^+} \frac{d}{d\beta} \Phi[(1-\beta)\xi + \beta\eta].$$

2. Si existe el gradiente de Φ entonces por su definición, podemos escribir:

$$\begin{aligned} F_{\Phi}(M_1, M_2) &= \text{Tr}\{[\nabla\Phi(M_1)](M_2 - M_1)\} \\ &= \langle \nabla\Phi(M_1), M_2 - M_1 \rangle. \end{aligned}$$

3. Si Φ es diferenciable en $M(\xi)$, llamando p al vector de pesos del diseño ξ y q al vector de pesos de un diseño η , por la relación entre la derivada direccional y la derivada de Gâteaux,

$$\begin{aligned} F_{\Phi}(\xi, \eta) &= F_{\Phi}(p, q) = G_{\Phi}(p, q - p) = (q - p)^t \frac{\partial\Phi}{\partial p} \\ &= \sum_{i=1}^k (q_i - p_i) \frac{\partial\Phi}{\partial p_i}. \end{aligned}$$

Denotando por $d_i = \frac{\partial\Phi}{\partial p_i}$, tenemos que

$$F_{\Phi}(p, e_j) = d_j - \sum_{i=1}^k p_i d_i.$$

Teorema 1.6. Si Φ es convexa en \mathcal{M} y diferenciable en $M(\xi^*)$, entonces

$$\xi^* \text{ es } \Phi\text{-óptimo si y sólo si } F_\Phi[M(\xi^*), f(x)f^t(x)] \geq 0$$

para todo $x \in \chi$.

Demostración.

La necesidad es inmediata haciendo uso del Teorema 1.4.

Veamos ahora que se verifica la condición de suficiencia.

Dada una matriz $M(\xi)$ sabemos que podemos expresarla de la forma:

$$M(\xi) = \sum_{i=1}^k \xi(x_i) f(x_i) f^t(x_i),$$

donde $\sum \xi(x_i) = 1$ y $\xi(x_i) > 0$. Entonces, si Φ es diferenciable en $M(\xi^*)$ (Lema 1.1.):

$$F_\Phi[M(\xi^*), M(\xi)] = \sum_i \xi(x_i) F_\Phi[M(\xi^*), f(x_i) f^t(x_i)].$$

Así, $F_\Phi[M(\xi^*), f(x) f^t(x)] \geq 0$ para todo $x \in \chi$ y esto implica que $F_\Phi[M(\xi^*), M(\xi)] \geq 0$ para todo $\xi \in \Xi$. Aplicando el Teorema anterior obtenemos el resultado.

□

Teorema 1.7. Si Φ es diferenciable en un entorno de $M(\xi^*)$, entonces son equivalentes:

1. ξ^* es Φ -óptimo
2. $f^t(x) \nabla \Phi[M(\xi^*)] f(x) \geq \text{Tr} M(\xi^*) \nabla \Phi[M(\xi^*)], x \in \chi$
3. $\inf_{x \in \chi} f^t(x) \nabla \Phi[M(\xi^*)] f(x) = \sum_{x \in \chi} f^t(x) \nabla \Phi[M(\xi^*)] f(x) \xi^*(x)$

La demostración puede encontrarse en Pázman [Paz86].

Observación 1.8. *De acuerdo con el segundo punto de la observación 1.7.*

$$F_{\Phi}[M(\xi^*), f(x)f^t(x)] = f^t(x)\nabla\Phi[M(\xi^*)]f(x) - \text{Tr}M(\xi^*)\nabla\Phi[M(\xi^*)].$$

El Teorema anterior proporciona un par de condiciones para la búsqueda del soporte óptimo

Corolario 1.2. *Bajo las condiciones del Teorema anterior, un diseño ξ^* es Φ -óptimo si y sólo si*

1. $F_{\Phi}[M(\xi^*), f(x_j)f^t(x_j)] = 0$ si $\xi^*(x_j) > 0$
2. $F_{\Phi}[M(\xi^*), f(x_j)f^t(x_j)] \geq 0$ si $\xi^*(x_j) = 0$.

1.6. Aproximación al diseño óptimo

En la mayoría de los problemas reales no es posible calcular analíticamente el diseño óptimo bajo un determinado criterio. Los procedimientos iterativos son herramientas muy útiles para la construcción de diseños aproximados. Una vez obtenidos estos, sería conveniente saber cuándo estamos lo “suficientemente” cerca del óptimo. Un indicador de esta distancia nos la proporciona la *eficiencia del diseño*, aunque en general esta cantidad tampoco la podremos conocer cuando el diseño óptimo no lo conozcamos, y nos tendremos que conformar con cotas inferiores de ella.

Definición 1.10. *Se dice que una función criterio Φ es homogénea si verifica $\Phi(\delta M) = \frac{1}{\delta}\Phi(M)$, siendo δ cualquier número real positivo.*

Definición 1.11. Se define la eficiencia de un diseño ξ respecto de una función criterio Φ homogénea como

$$eff_{\Phi}(\xi) = \frac{\min_{\mu} \Phi[M(\mu)]}{\Phi[M(\xi)]} = \frac{\Phi(\xi^*)}{\Phi(\xi)},$$

siendo ξ^* el diseño Φ - óptimo.

Observación 1.9.

1. $eff_{\Phi}(\xi) \in [0, 1]$. Cuanto más cerca esté del 1 mejor será la eficiencia. (A menudo la eficiencia se expresa en porcentajes).
2. Significado de la eficiencia. Supongamos que para ξ se ha tomado una muestra de tamaño N , es decir, se han realizado N experimentos, y que su eficiencia es ρ . ¿Qué tamaño de muestra se ha de tomar con ξ^* para obtener la misma precisión en las estimaciones? (Obviamente si ξ^* es el óptimo, se necesitarán menos observaciones).

$$\rho = \frac{N^* \Phi(M^*)}{N \Phi(M)} = \frac{N^*}{N} \cdot 1 \Rightarrow N^* = \rho \cdot N$$

Así por ejemplo, si la eficiencia de un diseño es del 50 %, entonces bastará tomar la mitad de observaciones con el diseño óptimo para obtener la misma precisión que con el diseño original.

3. Cota de la eficiencia.

De acuerdo con el Lema (1.1)

$$F_{\Phi}(M, N) \leq \Phi(N) - \Phi(M),$$

por tanto

$$\inf_N F_{\Phi}(M, N) \leq \Phi(M(\xi^*)) - \Phi(M),$$

y dividiendo por $\Phi(M)$ resulta:

$$eff_{\Phi}(\xi) \geq 1 + \frac{\inf_N F_{\Phi}(M, N)}{\Phi(M)}.$$

De este modo obtenemos una cota para la eficiencia de un diseño dado. Esta cota puede usarse como regla de parada para los algoritmos de construcción de diseños óptimos.

Además, si Φ es diferenciable, por el Lema 1.3., se tiene que:

$$\begin{aligned} eff_{\Phi} &\geq 1 + \frac{\inf_N F_{\Phi}(M, N)}{\Phi(M)} \\ &= 1 + \frac{\inf_{x \in \chi} F_{\Phi}[M(\xi), f(x)f^t(x)]}{\Phi(M)}. \end{aligned}$$

La idea propuesta por Wynn [Wyn70] y Fedorov [Fed72] para la construcción de medidas de diseño Φ -óptimas es la siguiente: supongamos que tenemos una medida de diseño $\xi^{(n)}$ con matriz de información $M(\xi^{(n)})$ y que Φ es diferenciable en $M(\xi^{(n)})$. Determinamos el punto x_{n+1} , que es aquél que maximiza $F_{\Phi}[M(\xi^{(n)}), f(x)f^t(x)]$ sobre χ . Si $\xi^{(n)}$ no es Φ -óptimo, entonces por el Teorema de equivalencia este valor máximo es menor que cero. Así, podemos aumentar Φ moviéndonos desde $M(\xi^{(n)})$ en la dirección de $f(x_{n+1})f^t(x_{n+1})$. Esto lo hacemos poniendo más peso en x_{n+1} que el que pone $\xi^{(n)}$. El nuevo diseño se obtiene de acuerdo a la regla:

$$\xi^{(n+1)} = (1 - \alpha_{n+1})\xi^{(n)} + \alpha_{n+1}\xi_{x_{n+1}},$$

siendo $\xi_{x_{n+1}}$ la medida con probabilidad 1 en x_{n+1} . En ocasiones denotamos a esta medida por $\mathbf{1}_{x_{n+1}}$.

Obsérvese que la matriz de información de $\xi^{(n+1)}$, viene dada por:

$$M(\xi^{(n+1)}) = (1 - \alpha_{n+1})M(\xi^{(n)}) + \alpha_{n+1}f(x_{n+1})f^t(x_{n+1}).$$

Ahora debemos elegir de manera adecuada la longitud del paso α_{n+1} , pues no todas las elecciones aseguran la convergencia y menos

aún que $\Phi[M(\xi^{(n+1)})] < \Phi[M(\xi^{(n)})]$. Obviamente, tomando α_{n+1} de una manera “adecuada”, podremos asegurar esta desigualdad, ya que

$$F_{\Phi}(M(\xi^{(n)}), f(x_{n+1})f^t(x_{n+1})) < \Phi[f(x_{n+1})f^t(x_{n+1})] - \Phi[M(\xi^{(n)})] < 0.$$

Una posibilidad es elegir α_{n+1} de manera que proporcione un aumento máximo en Φ y, suponiendo que Φ es diferenciable, la longitud de paso se determina por $F_{\Phi}[M(\xi^{(n+1)}), f(x_{n+1})f^t(x_{n+1})] = 0$.

Otra forma de elegir longitudes de paso para este tipo de algoritmos iterativos es fijar una sucesión $\{\alpha_{n+1}\}$ de longitudes de paso de antemano, sin preocuparnos si Φ decrece en cada iteración. La teoría general asegura que si hacemos esto, para acercarnos al óptimo $M(\xi^*)$, debemos requerir que $\alpha_{n+1} \rightarrow 0$ cuando $n \rightarrow \infty$. Además $\sum \alpha_{n+1}$ debe ser divergente.

Fedorov y Hackl [Fed97] proponen las siguientes tres elecciones. Cualquier sucesión $\{\alpha_{n+1}\}$ tal que:

$$\lim_{n \rightarrow \infty} \alpha_{n+1} = 0, \quad \sum_{n=0}^{\infty} \alpha_{n+1} = \infty; \quad (1.2)$$

$$\alpha_{n+1} = \arg \min_{\alpha} \Phi[(1 - \alpha)\xi^{(n)} + \alpha\xi(x_n)]; \quad (1.3)$$

$$\alpha_{n+1} = \begin{cases} \alpha_n, & \text{si } \Phi[(1 - \alpha_n)\xi^{(n)} + \alpha_n\xi(x_n)] < \Phi(\xi^{(n)}), \\ \gamma\alpha_n, & \text{en otro caso.} \end{cases} \quad (1.4)$$

La cantidad $\gamma < 1$ debe ser elegida de manera apropiada.

La regla de elección (1.3) se denomina regla del descenso más pronunciado (steepest descent).

1.6.1. Algoritmo de Wynn-Fedorov para D -optimización

Decíamos en la introducción que el problema de construcción de medidas de diseño Φ -óptimas no es tan sencillo si no podemos identificar inicialmente un conjunto finito que incluya los puntos de soporte de la medida óptima. Por esta razón se han desarrollado algoritmos especiales para este propósito. Por ejemplo, para el caso concreto del criterio de D -optimización tenemos el algoritmo de Wynn-Fedorov.

Debido a que en cada paso utilizaremos el diseño calculado anteriormente, es útil obtener la expresión explícita del determinante en un paso en función del calculado en la iteración anterior.

Teorema 1.8. *Sea $M(\xi^{(n)})$ la matriz de información de un diseño no degenerado $\xi^{(n)}$ y sea $M(\xi_{x_{n+1}})$ la matriz de información de un diseño concentrado en un único punto x_{n+1} . Entonces el determinante de la matriz de información de una combinación lineal convexa de estos diseños $\xi^{(n+1)} = (1 - \alpha_{n+1})\xi^{(n)} + \alpha_{n+1}\xi_{x_{n+1}}$ es igual a:*

$$|M(\xi^{(n+1)})| = (1 - \alpha_{n+1})^m \left(1 + \frac{\alpha_{n+1} \lambda(x_{n+1}) d(x_{n+1}, \xi^{(n)})}{1 - \alpha_{n+1}} \right) |M(\xi^{(n)})|. \quad (1.5)$$

Observación 1.10. *De (1.5) no es difícil concluir que el determinante de la matriz de información del diseño $\xi^{(n+1)}$ depende de la cantidad α_{n+1} y de las coordenadas del punto en el cual el diseño $\xi_{x_{n+1}}$ se concentra. Veremos que es siempre posible encontrar un punto x_{n+1} y un α_{n+1} para el cual $|M(\xi^{(n+1)})| > |M(\xi^{(n)})|$, si el diseño $\xi^{(n)}$ no es D -óptimo.*

Teorema 1.9. *El valor más grande posible de $|M(\xi^{(n+1)})|$ para un diseño dado $\xi^{(n)}$ es:*

$$\max_{x,\alpha} |M(\xi^{(n+1)})| = \frac{[\lambda(x_{n+1})d(x_{n+1}, \xi^{(n)})]^m (m-1)^{m-1}}{[\lambda(x_{n+1})d(x_{n+1}, \xi^{(n)}) - 1]^{m-1} m^m} |M(\xi^{(n)})| > |M(\xi^{(n)})|,$$

donde x_{n+1} es la solución de la ecuación

$$\lambda(x_{n+1})d(x_{n+1}, \xi^{(n)}) = \max_x \lambda(x)d(x, \xi^{(n)}),$$

y α_{n+1} viene definido como:

$$\alpha_{n+1} = \frac{\lambda(x_{n+1})d(x_{n+1}, \xi^{(n)}) - m}{(\lambda(x_{n+1})d(x_{n+1}, \xi^{(n)}) - 1)m}.$$

Una demostración de este Teorema puede encontrarse en Fedorov, [Fed72].

Teniendo en cuenta estos dos últimos teoremas, el algoritmo es el siguiente:

1. Tomamos un diseño inicial $\xi^{(0)}$ regular formado por n puntos x_1, \dots, x_n y pesos $\xi^{(0)}(x_1), \dots, \xi^{(0)}(x_n)$ respectivamente ($\sum_{i=1}^n \xi^{(0)}(x_i) = 1$, $n \geq m$).

2. Calculamos la matriz de información del diseño $\xi^{(0)}$,

$$M(\xi^{(0)}) = \sum_{i=1}^n \xi^{(0)}(x_i) \lambda(x_i) f(x_i) f^t(x_i)$$

y su inversa.

3. Se busca el punto x^1 donde se maximiza la función $\lambda(x)d(x, \xi^{(0)})$ y se calcula su valor en el punto x^1 : $\lambda(x^1)d(x^1, \xi^{(0)})$.
4. Se construye el diseño $\xi^{(1)} = (1-\alpha_1)\xi^{(0)} + \alpha_1 \xi_{x^1}$ que estará formado por los puntos y pesos siguientes:

$$\xi^{(1)} = \left\{ \begin{array}{ccccc} x_1 & x_2 & \dots & x_n & x^1 \\ (1-\alpha_1)p_1 & (1-\alpha_1)p_2 & \dots & (1-\alpha_1)p_n & \alpha_1 \end{array} \right\}$$

5. Se elige el valor α_1 que incremente al máximo el valor del determinante de la matriz de información del nuevo diseño $M(\xi^{(1)})$ que, según el Teorema 1.9, es:

$$\alpha_1 = \frac{\lambda(x^1)d(x^1, \xi^{(0)}) - m}{(\lambda(x^1)d(x^1, \xi^{(0)}) - 1)m}.$$

6. Se comprueba si se verifica alguna de las siguientes condiciones:

$$\alpha_1 < \gamma_1,$$

$$\frac{|M(\xi^{(1)})| - |M(\xi^{(0)})|}{|M(\xi^{(1)})|} < \gamma_2, \quad (1.6)$$

$$\frac{\lambda(x^1)d(x^1, \xi^{(0)}) - m}{m} < \gamma_3,$$

donde $\gamma_1, \gamma_2, \gamma_3$ son precisiones prefijadas.

7. Si se verifica alguna de las condiciones del paso anterior, el diseño $\xi^{(1)}$ está suficientemente próximo al D-óptimo.
Si no se verifica ninguna de las reglas de parada anteriores se repiten los pasos del 2 al 7 actualizando $\xi^{(0)}$ a $\xi^{(1)}$ y así sucesivamente.

Para la construcción iterativa de las matrices $M(\xi^{(0)}), M(\xi^{(1)}), \dots$ se utilizan los resultados de los teoremas anteriores.

El procedimiento iterativo anterior converge y su diseño límite $\tilde{\xi}$ coincide con uno de los diseños D-óptimos.

Teorema 1.10. *Supongamos que se satisfacen las condiciones del Teorema de equivalencia. Entonces el procedimiento iterativo 1-7 converge, en cuyo caso:*

$$\lim_{n \rightarrow \infty} |M(\xi^{(n)})| = |M(\xi^*)|,$$

donde $M(\xi^*)$ es la matriz de información correspondiente al diseño D-óptimo.

Ejemplo 1.1.

Sea $\chi = [-1, 1]$ y el modelo

$$E[y(x)] = \theta^t f(x) = \theta_1 + \theta_2 x, \quad \sigma^2(x) = 1, \quad x \in \chi.$$

Aplicaremos el algoritmo de Wynn-Fedorov para ver cómo mejora el diseño inicial. Tomemos como diseño inicial el siguiente:

$$\xi^{(0)} = \begin{Bmatrix} -1 & 0 \\ 0.5 & 0.5 \end{Bmatrix}$$

La sucesión de diseños aplicando los pasos del algoritmo en las diez primeras iteraciones, quedan resumidas en la Tabla 1.1.

Observamos que la sucesión de diseños tiende al diseño óptimo

$$\xi^* = \begin{Bmatrix} -1 & 1 \\ 0.5 & 0.5 \end{Bmatrix}$$

El algoritmo introduce de manera alternativa los puntos 1 y -1. El valor de la función de varianza generalizada decrece gradualmente, pero no monótonamente hacia 2.

Una característica del algoritmo es que el diseño al que tiende va incorporando la misma cantidad de puntos en -1 que en 1, que es lo que da lugar al óptimo. Por otro lado, la eficiencia aumenta rápidamente hacia uno. Solamente, después de añadir el primer punto al diseño, esta aumenta de 0.5 a 0.833.

Cota de Atwood

Se han considerado anteriormente cotas para la D-eficiencia de dicho diseño basadas en el Teorema de equivalencia. Dichas cotas

x^i	Diseño	$d(x, \xi_N)$	Eff
1	$\xi^{(1)} = \begin{Bmatrix} -1 & 0 & 1 \\ \frac{5}{18} & \frac{5}{18} & \frac{4}{9} \end{Bmatrix}$	10	0.5 %
-1	$\xi^{(2)} = \begin{Bmatrix} -1 & 0 & 1 \\ 0.450 & 0.209 & 0.335 \end{Bmatrix}$	2.960	0.833 %
1	$\xi^{(3)} = \begin{Bmatrix} -1 & 0 & 1 \\ 0.368 & 0.169 & 0.463 \end{Bmatrix}$	2.613	0.880 %
-1	$\xi^{(4)} = \begin{Bmatrix} -1 & 0 & 1 \\ 0.470 & 0.140 & 0.390 \end{Bmatrix}$	2.456	0.906 %
1	$\xi^{(5)} = \begin{Bmatrix} -1 & 0 & 1 \\ 0.410 & 0.120 & 0.470 \end{Bmatrix}$	2.365	0.922 %
-1	$\xi^{(6)} = \begin{Bmatrix} -1 & 0 & 1 \\ 0.470 & 0.110 & 0.420 \end{Bmatrix}$	2.305	0.933 %
1	$\xi^{(7)} = \begin{Bmatrix} -1 & 0 & 1 \\ 0.430 & 0.090 & 0.480 \end{Bmatrix}$	2.262	0.941 %
-1	$\xi^{(8)} = \begin{Bmatrix} -1 & 0 & 1 \\ 0.480 & 0.080 & 0.440 \end{Bmatrix}$	2.230	0.948 %
1	$\xi^{(9)} = \begin{Bmatrix} -1 & 0 & 1 \\ 0.438 & 0.080 & 0.481 \end{Bmatrix}$	2.205	0.953 %
-1	$\xi^{(10)} = \begin{Bmatrix} -1 & 0 & 1 \\ 0.484 & 0.070 & 0.445 \end{Bmatrix}$	2.185	0.957 %

Tabla: 1.1: Construcción de un diseño D –óptimo para el modelo de regresión lineal a través del origen.

pueden ser mejoradas para el criterio de D -optimización (Atwood [Atw69]):

Teorema 1.11. *Sea $M(\xi)$ una matriz no singular. Entonces*

$$eff_D(\xi) = \left(\frac{|M(\xi)|}{\max_{\xi} |M(\xi)|} \right)^{1/m} \geq \frac{m}{\max_x d(x, \xi)}. \quad (1.7)$$

Demostración:

La matriz de información, $M(\xi)$, es una matriz simétrica definida positiva, por tanto existe U ortogonal tal que $UM(\xi)U^t = I$.

Consideremos un modelo, cuyas funciones de regresión sean $\varphi(x) = Uf(x) \forall x \in \chi$. Llamando $d_g(x, \xi) = g^t(x)M^{-1}(\xi)g(x)$, se verifica que

$$d(x, \xi) = d_{\varphi}(x, \xi),$$

es decir, la función de dispersión es invariante por transformaciones lineales. Además dado ξ' un diseño cualquiera, se verifica que $\frac{\det M(\xi)}{\det M(\xi')} \propto \frac{\det M_{\varphi}(\xi)}{\det M_{\varphi}(\xi')}$, siendo $M_{\varphi}(\xi)$ la matriz del diseño para el modelo $\varphi(x)$. Por tanto, podemos considerar en el resto de la demostración que $M(\xi) = I$.

Sea ξ^* un diseño D -óptimo. Como

$$\int_{\chi} d(x, \xi) d\xi^*(x) \leq \int_{\chi} \max_x d(x, \xi) d\xi^*(x) = \max_x d(x, \xi)$$

resulta

$$\begin{aligned} \max_x d(x, \xi) &\geq \int_{\chi} d(x, \xi) d\xi^*(x) = \int_{\chi} f^t(x) M^{-1}(\xi) f(x) d\xi^*(x) \\ &= \int_{\chi} Tr[f^t(x) M^{-1}(\xi) f(x)] d\xi^*(x) \\ &= Tr \left(M^{-1}(\xi) \int_{\chi} f^t(x) f(x) d\xi^*(x) \right) \\ &= Tr(I \cdot M(\xi^*)) = Tr(M(\xi^*)). \end{aligned}$$

De modo que usando la desigualdad de la media aritmética y geométrica, obtenemos que

$$\left(\frac{|M(\xi)|}{|M(\xi^*)|} \right)^{1/m} \geq \frac{m}{Tr[M(\xi^*)]} \geq \frac{m}{\max_x d(x, \xi)}$$

esto es,

$$eff_D(\xi) \geq \frac{m}{\max_x d(x, \xi)}.$$

□

Considerando la función criterio $\Phi_D[M(\xi)] = |M(\xi)|^{-1/m}$, resulta:

$$\nabla(\log \Phi_D(\xi)) = \frac{1}{m} \nabla(-\log |M(\xi)|) = -\frac{1}{m} M^{-1}(\xi).$$

Por otra parte como

$$\nabla(\log \Phi_D(\xi)) = \frac{\nabla \Phi_D(\xi)}{\Phi_D(\xi)}.$$

De ambas igualdades se deduce que:

$$\nabla \Phi_D(\xi) = \frac{-1}{m|M(\xi)|^{1/m}} M^{-1}(\xi).$$

Teniendo en cuenta lo anterior,

$$\begin{aligned} F_{\Phi_D}[M(\xi), f(x)f^t(x)] &= Tr\{[\nabla(M(\xi))][f(x)f^t(x) - M(\xi)]\} \\ &= \frac{1}{m|M(\xi)|^{1/m}} \left\{ m - f^t(x)M^{-1}(\xi)f(x) \right\} \\ &= |M(\xi)|^{-1/m} \left\{ 1 - \frac{f^t(x)M^{-1}(\xi)f(x)}{m} \right\}. \end{aligned}$$

De este modo, teniendo en cuenta que Φ_D es diferenciable y que por resultados anteriores habíamos obtenido que la cota para la eficiencia era

$$eff_{f\Phi} \geq 1 + \frac{\inf_{x \in \chi} F_{\Phi}(M(\xi), f(x)f^t(x))}{\Phi(M)},$$

resulta que para D -optimización la cota para la eficiencia se puede escribir como:

$$\begin{aligned} eff_D(\xi) &\geq 1 + \frac{|M(\xi)|^{-1/m} \left\{ 1 - \frac{f^t(x)M^{-1}(\xi)f(x)}{m} \right\}}{|M(\xi)|^{-1/m}} \\ &= 2 - \frac{\max_x d(x, \xi)}{m} \end{aligned} \quad (1.8)$$

Representando ambas cotas para la eficiencia, (1.7) y (1.8), podemos observar cuál es la mejor sobre el diseño óptimo, como podemos ver en la Figura 1.1. Además lo podemos demostrar mediante la siguiente:

Proposición 1.10. *La cota de la eficiencia proporcionada por Atwood [Atw69] mejora los resultados obtenidos por Kiefer [Kie60].*

Demostración.

$$2 - \frac{\max_x d(x, \xi)}{m} \leq \frac{m}{\max_x d(x, \xi)},$$

es decir,

$$2m \max_x d(x, \xi) - (\max_x d(x, \xi))^2 \leq m^2 \Leftrightarrow 0 \leq (m - \max_x d(x, \xi))^2.$$

□

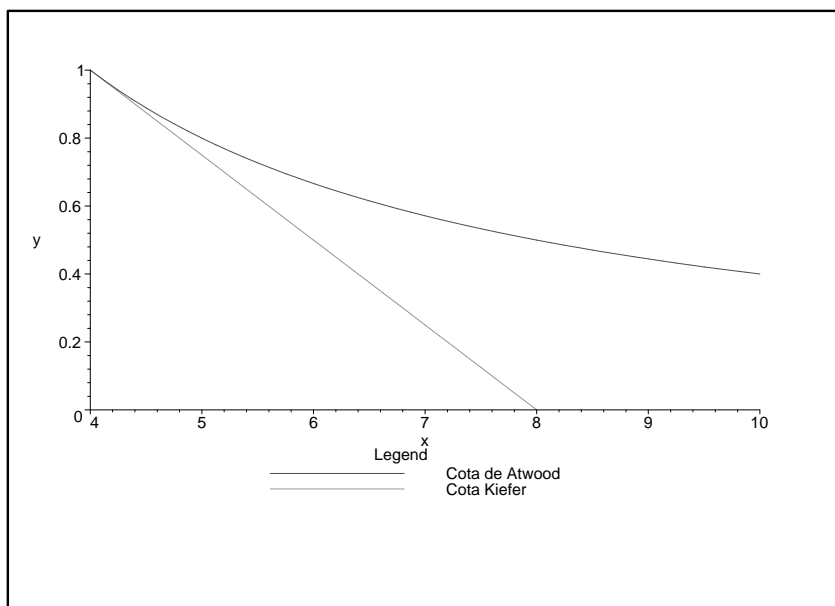


Figura 1.1 : Comparación de las cotas de Kiefer y Atwood.

Así, en el ejemplo anterior, el diseño del que partíamos

$$\xi^{(0)} = \begin{Bmatrix} -1 & 0 \\ 1/2 & 1/2 \end{Bmatrix}$$

tiene como cotas para la eficiencia:

Cota de Atwood

$$eff_D(\xi^{(0)}) \geq \frac{m}{\max_x d(x, \xi)} = 0.5.$$

Cota de Kiefer

$$eff_D(\xi^{(0)}) \geq 2 - \frac{\max_x d(x, \xi)}{m} = 0.$$

1.7. Modelos no lineales

Como indicábamos en la introducción del capítulo, la teoría de diseños óptimos se desarrolla en principio para modelos lineales. Sin embargo, en la práctica es frecuente encontrarnos con modelos no lineales.

En estos casos, los diseños óptimos dependen del valor de θ , es por esto que se denominan diseños **localmente óptimos**. Ahora no podemos definir la matriz de información como lo hemos hecho anteriormente. Utilizaremos la fórmula de Taylor para transformar el modelo no lineal en uno lineal. Para ello, es necesario hacer una estimación inicial de los parámetros, $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_m^{(0)})^t$. De modo que el modelo queda linealizado del siguiente modo:

$$\begin{aligned} E[y(x)] = \eta(x, \theta) &= \eta(x, \theta^{(0)}) + \sum_{i=1}^m \left(\frac{\partial \eta(x, \theta)}{\partial \theta_i} \right)_{\theta^{(0)}} (\theta_i - \theta_i^{(0)}) \\ &+ \frac{1}{2} \sum_{i,j=1}^m \left(\frac{\partial^2 \eta(x, \theta)}{\partial \theta_i \partial \theta_j} \right)_{\theta^{(x)}} (\theta_i - \theta_i^{(0)}) (\theta_j - \theta_j^{(0)}), \end{aligned}$$

donde $\theta_i^{(s)}$ está comprendido entre $\theta_i^{(0)}$ y θ_i , $i = 1, \dots, m$. Escribiendo:

$$\tilde{f}_i(x) = \left(\frac{\partial \eta(x, \theta)}{\partial \theta_i} \right)_{\theta^{(0)}}, \quad i = 1, \dots, m,$$

linealizamos el modelo, y se calcula el diseño óptimo para el modelo linealizado.

Capítulo 2

Diseños con variables no controlables

2.1. Introducción

Es frecuente, que en el diseño de un experimento se consideren todas las variables explicativas bajo el control del experimentador. Esta suposición frecuentemente no es realista. De hecho, en muchas áreas experimentales, es habitual encontrar variables que no están bajo control. Por ejemplo, en un experimento de tipo médico la asignación de tratamientos a los pacientes puede ser realizada por el experimentador. Sin embargo, hay otras variables que escapan a su control, como pueden ser la edad o la presión sanguínea. Si los valores concretos de estas variables son conocidos antes de realizar el experimento, entonces el objetivo será encontrar diseños óptimos condicionados a los valores conocidos de estas variables. Llamaremos a estos diseños marginalmente restringidos (MR).

Pueden existir otras variables cuyos valores concretos no pueden conocerse hasta finalizado el experimento. En esta situación puede ser factible suponer a priori una distribución de probabilidad condicionada, para posteriormente buscar la distribución marginal ópti-

ma relativa a la variable controlable. Llamaremos a estos diseños condicionalmente restringidos (CR).

Por otro lado, podemos encontrar problemas donde aparezcan los tres tipos de variables que hemos considerado: controlables, no controlables con valores conocidos de antemano, y no controlables con valores desconocidos antes de realizar el experimento. Llamaremos a estos diseños marginal y condicionalmente restringidos (MCR).

Ardanuy y López Fidalgo [Ard92] introducen la teoría de los diseños óptimos con soporte restringido. Cook y Thibodeau [Coo80] introducen la idea de diseño D –óptimo marginalmente restringido. Calvete [Cal83] realiza un estudio más completo sobre los diseños marginalmente restringidos. Nachtsheim [Nac89] proporciona teoremas de equivalencia y algoritmos para D –optimización y D_s –optimización. Huang y Hsu [Hua93] generalizan estos resultados para una función criterio general. Huang y Chang [Hua95] consideran el problema de los diseños óptimos marginalmente restringidos para modelos lineales. Cook y Fedorov [Coo95] tratan el problema general de diseños óptimos con restricciones en sentido general, ya sean sobre los parámetros, el soporte o las funciones. Schwabe y Wong [Sch04] plantean la relación existente entre las eficiencias de los diseños marginales y los diseños productos. Garcet-Rodríguez [Gar02] y López-Fidalgo y Garcet-Rodríguez [Lóp04a] prueban la convergencia de los algoritmos para criterios generales y proporcionan las herramientas necesarias para el tratamiento de diseños óptimos CR y MCR.

En este capítulo, realizamos una revisión del problema de construcción de diseños MR, CR y MCR, indicando los resultados básicos, algoritmos de construcción y una “nueva” cota de la eficiencia para cada uno de ellos basada en la cota de Atwood [Atw69].

2.2. Diseños Marginalmente Restringidos

Sea $y(x)$ la variable aleatoria observable, y supongamos que algunas de las variables de las que depende no pueden ser completamente controladas por el experimentador. Consideremos el espacio de diseño o dominio experimental $\chi = \chi_1 \times \chi_2$, donde χ_1 representa el espacio para las variables que por estar restringidas de algún modo, el diseño no podrá controlar. Cada punto $x \in \chi$ lo denotaremos por $x = (x_1, x_2)$.

Definición 2.1. Sea $\xi = \xi(x_1, x_2)$ un diseño arbitrario, llamamos **diseños marginales** a los diseños

$$\xi_i(x_i) = \int_{\chi_j} \xi(dx_j, x_i), \quad i \neq j = 1, 2. \quad (2.1)$$

Denotemos por Ξ el conjunto de todos los diseños en este espacio producto. Dado un diseño conjunto ξ , los diseños marginales para χ_1 y χ_2 serán las correspondientes distribuciones marginales ξ_1 y ξ_2 como hemos visto.

Si los valores de la variable x_1 son conocidos de antemano, estos determinan un diseño específico $\tilde{\xi}_1$ con **soporte finito**

$$S_{\tilde{\xi}_1} = \chi_1 = \{x_1^{(1)}, x_1^{(2)}, \dots, x_1^{(k)}\}.$$

De este modo, definimos el espacio Ξ_R como aquél que se restringe a aquellos diseños que tienen por diseño marginal $\tilde{\xi}_1$:

$$\Xi_R = \{\xi \in \Xi / \xi_1 = \tilde{\xi}_1\},$$

y denotamos por:

$$\mathcal{M}_R = \{M(\xi) \in \mathcal{M} / \xi \in \Xi_R\}.$$

Por la linealidad de la integral es fácil ver que se verifica que Ξ_R y \mathcal{M}_R **son convexos**.

Definición 2.2. Para cada diseño $\xi \in \Xi_R$ llamaremos a

$$\xi_{2|1}(x_2|x_1) \quad (2.2)$$

diseño **condicionado** asociado a ξ , sobre χ_2 dado $x_1 \in \chi_1$.

Teniendo en cuenta las definiciones anteriores, podemos escribir

$$\xi(x_1, x_2) = \xi_{2|1}(x_2|x_1) \cdot \tilde{\xi}_1(x_1).$$

También podemos decir que para cada $x_1^{(i)}$ fijo,

$$\xi \in \Xi_R \Leftrightarrow \sum_{x_2} \xi(x_1^{(i)}, x_2) = \tilde{\xi}_1(x_1^{(i)}) > 0.$$

Por tanto, en el sumatorio para cada $x_1^{(i)}$ y para algún x_2 , $\xi(x_1^{(i)}, x_2) > 0$.

De lo anterior, podemos deducir que si el número de puntos del soporte del diseño marginal es k , $\#S_{\tilde{\xi}_1} = k$, digamos,

$$S_{\tilde{\xi}_1} = \chi_1 = \{x_1^{(1)}, x_1^{(2)}, \dots, x_1^{(k)}\},$$

entonces:

$$\begin{aligned} \sum_{x_2} \xi(x_1^{(1)}, x_2) &= \tilde{\xi}_1(x_1^{(1)}) > 0 \\ \vdots &\quad \quad \quad \vdots \\ \sum_{x_2} \xi(x_1^{(k)}, x_2) &= \tilde{\xi}_1(x_1^{(k)}) > 0, \end{aligned}$$

y por tanto existen al menos k puntos de χ_2 tales que $\xi(x_1, x_2) > 0$.

Esto nos indica que no son posibles diseños unipuntuales, es decir, diseños de la forma:

$$\xi = \left\{ \begin{array}{c} (x_1, x_2) \\ \xi(x_1, x_2) \end{array} \right\},$$

de hecho, los diseños han de tener al menos k puntos.

Por tanto, las matrices de información de los diseños:

$$\xi = \left\{ \begin{array}{cccc} (x_1^{(1)}, x_2^{(1)}) & (x_1^{(2)}, x_2^{(2)}) & \dots & (x_1^{(k)}, x_2^{(k)}) \\ \tilde{\xi}_1(x_1^{(1)}) & \tilde{\xi}_1(x_1^{(2)}) & \dots & \tilde{\xi}_1(x_1^{(k)}) \end{array} \right\},$$

$$x_2^{(1)}, x_2^{(2)}, \dots, x_2^{(k)} \in \chi_2$$

son los generadores de \mathcal{M}_R , que es un conjunto convexo en un espacio euclídeo de dimensión $\frac{m(m+1)}{2}$, siendo m el número de parámetros del modelo.

El teorema de Caratheodory garantiza que cualquier matriz de información es una combinación convexa de $1 + \frac{m(m+1)}{2}$ generadores como máximo. Como $\xi = \xi_{2|1} \cdot \xi_1 = \xi_{2|1} \cdot \tilde{\xi}_1$ y el soporte de $\tilde{\xi}_1$ tiene k puntos y el de $\xi_{2|1}$ tiene como máximo $1 + \frac{m(m+1)}{2}$, entonces para cualquier matriz de información siempre existe un diseño conjunto asociado con no más de $k(1 + \frac{m(m+1)}{2})$ puntos en su soporte.

En particular, *el soporte del diseño condicional $\xi_{2|1}(\cdot|x_1)$ (que es el que se lleva a cabo en la práctica para cada $x_1 \in \chi_1$) tendrá como máximo $1 + \frac{m(m+1)}{2}$ puntos en su soporte para cada $x_1 \in \chi_1$.*

Por otro lado, si el soporte del diseño de $\tilde{\xi}_1$ es infinito, no hay garantía de que el diseño conjunto tenga soporte finito, pero el diseño condicional $\xi_{2|1}(\cdot|x_1)$ tendrá más de $1 + \frac{m(m+1)}{2}$ puntos en su soporte para un valor específico x_1 .

A continuación se presentan los principales resultados para los diseños marginalmente restringidos, que son muy similares a los que nos encontramos en la teoría de los diseños no restringidos que vimos en el capítulo primero.

Teorema 2.1. *Sea un diseño arbitrario $\xi \in \Xi_R$, entonces:*

1. *La matriz de información $M(\xi)$ es simétrica semidefinida positiva.*
2. *Si el soporte del diseño contiene menos de m puntos entonces $M(\xi)$ es singular.*

3. El conjunto \mathcal{M}_R es un subconjunto convexo de \mathcal{M} .
4. Toda matriz de información se puede escribir como

$$M(\xi) = \sum_{i=1}^k p_i f(x_i) f'(x_i),$$

donde

$$p_i \geq 0, \quad \sum_{i=1}^k p_i = 1, \quad k \leq \frac{m(m+1)}{2} + 1.$$

Observación 2.1. Notemos que como estamos suponiendo que el soporte de S_{ξ_1} es finito, podemos restringir Ξ_R a los diseños con soporte finito.

Sea Φ una función criterio convexa.

Definición 2.3. Sea $\xi^* \in \Xi_R$, diremos que un diseño es Φ -óptimo marginalmente restringido si verifica

$$\Phi[M(\xi^*)] = \min_{\xi \in \Xi_R} \Phi[M(\xi)]. \quad (2.3)$$

Teorema 2.2. El conjunto

$$\Xi_R^* = \{\xi \in \Xi_R \mid \xi \text{ es } \Phi - \text{óptimo MR}\}$$

es un subconjunto convexo de Ξ_R .

Lema 2.1. Si Φ es convexa, entonces ξ^* es Φ -óptimo marginalmente restringido si y solo si

$$F_\Phi[M(\xi^*), M(\xi)] \geq 0, \quad \forall \xi \in \Xi_R.$$

Teorema 2.3. (Teorema de equivalencia para diseños marginalmente restringidos.) Bajo la condición de convexidad de Φ , un diseño ξ^* Φ -óptimo marginalmente restringido puede caracterizarse equivalentemente por una las siguientes condiciones:

1. $\Phi[M(\xi^*)] = \inf_{\xi \in \Xi_R} \Phi[M(\xi)].$
2. $\inf_{N \in \mathcal{M}_R} F_\Phi[M(\xi^*), N] = \sup_{\xi \in \Xi_R^+} \inf_{N \in \mathcal{M}_R} F_\Phi[M(\xi), N].$
3. $\inf_{N \in \mathcal{M}_R} F_\Phi[M(\xi^*), N] = 0.$

En el caso de que la función criterio sea diferenciable,

Lema 2.2. *Sea $\xi \in \Xi_R^+$. Si la función criterio Φ es diferenciable en \mathcal{M}^+ , se verifica:*

$$\begin{aligned} \inf_{\eta \in \Xi_R} F_\Phi[M(\xi), M(\eta)] \\ = \int_{\chi_1} \inf_{x_2 \in \chi_2} F_\Phi[M(\xi), M_{(x_1, x_2)}] \tilde{\xi}_1(dx_1). \end{aligned}$$

Teorema 2.4. *Si la función Φ es diferenciable sobre \mathcal{M}^+ son equivalentes las siguientes afirmaciones:*

1. $\xi^* \in \Xi_R^*$
2. $\int_{\chi_1} \inf_{x_2 \in \chi_2} F_\Phi[M(\xi^*), M_{(x_1, x_2)}] \tilde{\xi}_1(dx_1) = \sup_{\xi \in \Xi_R^+} \int_{\chi_1} \inf_{x_2 \in \chi_2} F_\Phi[M(\xi), M_{(x_1, x_2)}] \tilde{\xi}_1(dx_1)$
3. $\int_{\chi_1} \inf_{x_2 \in \chi_2} F_\Phi[M(\xi^*), M_{(x_1, x_2)}] \tilde{\xi}_1(dx_1) = 0,$

donde $M_{(x_1, x_2)} = f(x_1, x_2)f^t(x_1, x_2)$ es la matriz de información del diseño $\xi \in \Xi$ concentrado en el punto $\mathbf{x} = (x_1, x_2)$.

Corolario 2.1. *Bajo las hipótesis del teorema anterior, si $\xi^* \in \Xi_R^*$ entonces*

$$\int_{\chi_1} \inf_{x_2 \in S_{\xi_2^*}} F_{\Phi}[M(\xi^*), M_{(x_1, x_2)}] \tilde{\xi}_1(dx_1) = 0,$$

siendo $S_{\xi_2^*}$ el soporte del diseño marginal ξ_2^* .

Demostración.

$$\begin{aligned} 0 &= F_{\Phi}[M(\xi^*), M(\xi^*)] = \\ &= \int_{\chi} F_{\Phi}[M(\xi^*), M_{(x_1, x_2)}] d\xi^*(x_1, x_2) = \\ &= \int_{\chi_1} \int_{\chi_2} F_{\Phi}[M(\xi^*), M_{(x_1, x_2)}] \xi_{2|1}^*(dx_2|x_1) \tilde{\xi}_1(dx_1) \geq \\ &\geq \int_{\chi_1} \inf_{x_2 \in S_{\xi_2}} F_{\Phi}[M(\xi^*), M_{(x_1, x_2)}] \tilde{\xi}_1(dx_1) \geq \\ &\geq \int_{\chi_1} \inf_{x_2 \in \chi_2} F_{\Phi}[M(\xi^*), M_{(x_1, x_2)}] \tilde{\xi}_1(dx_1) \\ &= 0, \text{ por ser } \xi^* \in \Xi_R^* \end{aligned}$$

□

Teorema 2.5. (Teorema de caracterización.) *Sea Φ diferenciable sobre \mathcal{M}^+ . Un diseño ξ^* es Φ -óptimo MR si y solo si $\forall x_1 \in \chi_1$*

$$\inf_{x_2 \in \chi_2} \partial \Phi[M(\xi^*), M_{(x_1, x_2)}] = \int_{\chi_2} \partial \Phi[M(\xi^*), M_{(x_1, x_2)}] \xi_{2|1}^*(dx_2|x_1).$$

Cotas para la eficiencia.

Como vimos en el capítulo anterior

$$F_{\Phi}[M(\xi), M(\eta)] \leq \Phi[M(\eta)] - \Phi[M(\xi)].$$

Por tanto, tomando el extremo inferior de η en el conjunto de diseños restringidos Ξ_R , resulta:

$$eff_{\Phi-MR}(\xi) \geq 1 + \frac{\inf_{\eta \in \Xi_R} F_{\Phi}[M(\xi), M(\eta)]}{\Phi[M(\xi)]}.$$

Por otro lado, en el caso en que Φ sea diferenciable, sabemos que la derivada direccional es lineal, por tanto:

$$\begin{aligned} & eff_{\Phi-MR}(\xi) \\ & \geq 1 + \frac{\inf_{\eta \in \Xi_R} F_{\Phi}[M(\xi), M(\eta)]}{\Phi[M(\xi)]} \\ & = 1 + \frac{\int_{\chi_1} \inf_{x_2 \in \chi_2} F_{\Phi}[M(\xi), f(x_1, x_2) f^t(x_1, x_2)] \eta_1(dx_1)}{\Phi[M(\xi)]} \\ & = 1 + \frac{\int_{\chi_1} \inf_{x_2 \in \chi_2} \{f^t(x_1, x_2) \nabla \Phi[M(\xi)] f(x_1, x_2) - Tr[M(\xi) \nabla \Phi[M(\xi)]]\} \tilde{\xi}_1(dx_1)}{\Phi[M(\xi)]}. \end{aligned}$$

De nuevo las cotas se usarán como regla de parada para algoritmos de construcción de diseños óptimos.

Teniendo en cuenta el Teorema 2.5 y la cota de la eficiencia, se propone el siguiente algoritmo:

1. Se establece un diseño inicial $\xi^{(0)} \in \Xi_R^+$.
2. A partir de $\xi^{(n)}$ calculamos el punto $x^{(n)} = (x_1^{(n)}, x_2^{(n)})$.
Tomando $\chi_1 = S_{\xi_1}^{\sim}$, $x^{(n)} = (x_1^{(n)}, x_2^{(n)}) =$

$$\begin{aligned} & arg \left\{ \inf_{x_1 \in \chi_1} \left\{ \inf_{x_2 \in \chi_2} F_{\Phi}[M(\xi^{(n)}), M_{(x_1, x_2)}] - \right. \right. \\ & \left. \left. \int_{\chi_2} F_{\Phi}[M(\xi^{(n)}), M_{(x_1, x_2)}] \xi_{2|1}^{(n)}(dx_2|x_1) \right\} \right\} \end{aligned}$$

3. Calculamos el diseño condicional (en $x^{(n)}$)

$$\xi_{2|1}^{(n+1)}(x_2|x_1^{(n)}) = (1 - \alpha_n) \xi_{2|1}^{(n)}(x_2|x_1^{(n)}) + \alpha_n 1_{(x_1^{(n)}, x_2^{(n)})}$$

4. Finalmente, construimos el diseño conjunto:

$$\xi^{(n+1)}(x_1, x_2) = \begin{cases} \xi^{(n)}(x_1, x_2) = \xi_{2|1}^{(n)}(x_2|x_1) \cdot \tilde{\xi}_1(x_1) & \text{si } x_1 \neq x_1^{(n)} \\ \xi_{2|1}^{(n+1)}(x_2|x_1^{(n)}) \cdot \tilde{\xi}_1(x_1^{(n)}) & \text{si } x_1 = x_1^{(n)} \end{cases}$$

Por tanto:

$$\begin{aligned} \xi^{(n+1)}(x_1, x_2) &= \xi_{2|1}^{(n)}(x_2|x_1) \cdot \tilde{\xi}_1(x_1) - \xi_{2|1}^{(n)}(x_2|x_1^{(n)}) \cdot \tilde{\xi}_1(x_1^{(n)}) \cdot \mathbf{1}_{x_1=x_1^{(n)}} \\ &\quad + \xi_{2|1}^{(n+1)}(x_2|x_1^{(n)}) \cdot \tilde{\xi}_1(x_1^{(n)}) \cdot \mathbf{1}_{x_1=x_1^{(n)}} \\ &= \xi^{(n)}(x_1, x_2) \\ &\quad + \alpha_n \tilde{\xi}_1(x_1^{(n)}) \mathbf{1}_{x_1=x_1^{(n)}} (\xi_{(x_1^{(n)}, x_2^{(n)})} - \xi_{2|1}^{(n)}(x_2|x_1^{(n)})) \end{aligned}$$

donde

$$\mathbf{1}_{x_1=x_1^{(n)}} = \begin{cases} 1 & \text{si } x_1 = x_1^{(n)} \\ 0 & \text{si } x_1 \neq x_1^{(n)} \end{cases}$$

5. El algoritmo se detiene cuando

$$1 + \frac{\int_{\chi_1} \inf_{x_2 \in \chi_2} \partial \Phi[M(\xi^{(n)}), M_{(x_1, x_2)}] \tilde{\xi}_1(dx_1)}{\Phi[M(\xi^{(n)})]} \geq \delta,$$

siendo $0 < \delta < 1$ es una cota inferior para la eficiencia.

Puede demostrarse (véase López Fidalgo y Garcet Rodríguez [Lóp04a]) que

$$\lim_{n \rightarrow \infty} \Phi[M(\xi^{(n)})] = \Phi[M(\xi^*)],$$

cuando la función criterio, Φ , cumple las siguientes condiciones:

1. Φ está acotada inferiormente y alcanza su mínimo en $M(\xi^*)$,
2. Φ tiene las derivadas de segundo orden acotadas, es decir, existe $B > 0$, tal que

$$|\partial^2 \Phi / \partial M_{ij} \partial M_{rt}| \leq B < \infty,$$

3. $0 < \alpha_n < 1$, $\lim_{n \rightarrow \infty} \alpha_n = 0$, $\sum_{n=0}^{\infty} \alpha_n = \infty$

$$\text{y } \sum_{n=0}^{\infty} \alpha_n^2 \leq k < \infty.$$

2.2.1. Criterios de optimización clásicos para diseños MR

En este trabajo se hace una aplicación de los resultados anteriores a dos de los criterios de optimalidad más usuales de la literatura: D -optimización, A -optimización, particularizando las propiedades generales.

D-optimización

Sea

$$\Phi_D[M(\xi)] = -\log |M(\xi)|. \quad (2.4)$$

La minimización de esta función es equivalente al criterio D -optimización para diseños marginalmente restringidos que requiere la minimización de $|M^{-1}(\xi)|$. Ya dijimos que esta función es estrictamente convexa sobre \mathcal{M} y diferenciable sobre \mathcal{M}^+ .

Además, mediante el siguiente lema demostramos cuál es el valor de su derivada direccional:

Lema 2.3. Sean $M_1 \in \mathcal{M}^+$, $M_2 \in \mathcal{M}$, entonces:

$$F_{\Phi_D}[M_1, M_2] = -[Tr(M_1^{-1}M_2) - m] \quad (2.5)$$

Demostración.

Obsérvese que por ser Φ diferenciable:

$$\begin{aligned} F_{\Phi_D}[M_1, M_2] &= \lim_{\alpha \rightarrow 0} \frac{\Phi_D[M_1 + \alpha(M_2 - M_1)] - \Phi_D[M_1]}{\alpha} \\ &= \left. \frac{\partial}{\partial \alpha} \Phi_D[M_1 + \alpha(M_2 - M_1)] \right|_{\alpha=0^+}. \end{aligned}$$

Como $\frac{\partial}{\partial \alpha} \ln|A| = Tr \left[A^{-1} \left(\frac{\partial A}{\partial \alpha} \right) \right]$, entonces:

$$\begin{aligned}
F_{\Phi_D}[M_1, M_2] &= \frac{\partial}{\partial \alpha} - \ln | [M_1 + \alpha(M_2 - M_1)] | \Big|_{\alpha=0^+} \\
&= - \left[\text{Tr}([M_1 + \alpha(M_2 - M_1)]^{-1}(M_2 - M_1)) \right] \Big|_{\alpha=0^+} \\
&= - [\text{Tr}(M_1^{-1}M_2) - m].
\end{aligned}$$

□

Haciendo uso del lema anterior y del teorema de equivalencia para diseños MR, resulta:

Teorema 2.6. (Teorema de equivalencia)

Dada la función Φ_D definida en (2.4), por ser Φ diferenciable en \mathcal{M}^+ , son equivalentes las siguientes proposiciones:

1. ξ^* es D -óptimo marginalmente restringido, $\xi^* \in \Xi_R^*$.
2. Siendo $d(x_1, x_2, \xi)$ la función de dispersión,

$$\int_{\chi_1} \max_{x_2 \in \chi_2} d(x_1, x_2, \xi^*) \tilde{\xi}_1(dx_1) = \min_{\xi \in \Xi_R^+} \int_{\chi_1} \max_{x_2 \in \chi_2} d(x_1, x_2, \xi) \tilde{\xi}_1(dx_1).$$

$$3. \int_{\chi_1} \max_{x_2 \in \chi_2} d(x_1, x_2, \xi^*) \tilde{\xi}_1(dx_1) = m.$$

Demostración.

Basta tener en cuenta que por el lema anterior y por ser $\text{Tr}[M_x M^{-1}(\xi)] = f^t(x) M^{-1}(\xi) f(x)$, se verifica que:

$$\begin{aligned}
F_{\Phi_D}[M(\xi), M_{(x_1, x_2)}] &= - [\text{Tr}(M(\xi)^{-1} M_{(x_1, x_2)}) - m] \\
&= - [f^t(x_1, x_2) M^{-1}(\xi) f(x) - m].
\end{aligned}$$

De modo que debido al Teorema 2.4 obtenemos el resultado.

□

Corolario 2.2. *Sea ξ^* el diseño D -óptimo marginalmente restringido, entonces:*

$$\int_{\chi_1} \max_{x_2 \in S_{\xi_2}} d(x_1, x_2, \xi^*) \tilde{\xi}_1(dx_1) = m.$$

Teorema 2.7. *El diseño ξ^* es D -óptimo marginalmente restringido si y solo si $\forall x_1 \in \chi_1$*

$$\max_{x \in \chi_2} d(x_1, x_2, \xi^*) = \int_{\chi_2} d(x_1, x_2, \xi^*) \xi_{2|1}^*(dx_2|x_1).$$

Cota de la eficiencia

Para el caso restringido, la cota para la eficiencia, viene dada por:

$$eff_{D-MR}(\xi) \geq 1 + \frac{\inf_G \text{generadores de } M_R F_{\Phi_D}[M(\xi), G]}{\Phi_D[M(\xi)]}$$

y por el teorema de equivalencia, para funciones diferenciables

$$\begin{aligned} eff_{D-MR}(\xi) &\geq 1 + \frac{\int_{\chi_1} \inf_{x_2} F_{\Phi_D}[M(\xi), f(x_1, x_2) f^t(x_1, x_2)]}{\Phi_D[M(\xi)]} \\ &= 1 + \frac{\int_{\chi_1} \tilde{\xi}_1(x_1) \left\{ |M(\xi)|^{-1/m} \left[1 - \frac{\max_{x_2} f^t(x_1, x_2) M^{-1}(\xi) f(x_1, x_2)}{m} \right] \right\}}{|M(\xi)|^{-1/m}} \\ &= 2 - \frac{1}{m} \int_{\chi_1} \tilde{\xi}_1(x_1) \max_{x_2} f^t(x_1, x_2) M^{-1}(\xi) f(x_1, x_2). \end{aligned}$$

Algoritmo de construcción de diseños D-óptimos marginalmente restringidos

Basándonos en el teorema de equivalencia y los teoremas de caracterización de diseños D -óptimos MR, presentamos el siguiente algoritmo:

1. Establecemos un diseño inicial $\xi^{(0)} \in \Xi_R^+$.
2. A partir de $\xi^{(n)}$ calculamos el punto $x^{(n)} = (x_1^{(n)}, x_2^{(n)})$ de $\chi_1 \times \chi_2$ (recuérdese que tomamos $\chi_1 = S_{\xi_1}^-$), donde en este punto se alcanza el máximo:

$$\begin{aligned} \max_{x_1 \in \chi_1 = S_{\xi_1}^-} \left\{ \max_{x_2 \in \chi_2} d(x_1, x_2, \xi^{(n)}) - \int_{\chi_2} d(x_1, x_2, \xi^{(n)}) \xi_{2|1}^{(n)}(dx_2|x_1) \right\} = \\ = d(x_1^{(n)}, x_2^{(n)}, \xi^{(n)}) - \int_{\chi_2} d(x_1^{(n)}, x_2^{(n)}, \xi^{(n)}) \xi_{2|1}^{(n)}(dx_2|x_1^{(n)}) \end{aligned}$$

3. Calculamos el diseño condicional (en el nuevo punto calculado, en los otros no varía)

$$\xi_{2|1}^{(n+1)}(x_2|x_1^{(n)}) = (1 - \alpha_n) \xi_{2|1}^{(n)}(x_2|x_1^{(n)}) + \alpha_n \mathbf{1}_{(x_1^{(n)}, x_2^{(n)})}$$

4. Finalmente, construimos el diseño conjunto:

$$\xi^{(n+1)}(x_1, x_2) = \begin{cases} \xi^{(n)}(x_1, x_2) = \xi_{2|1}^{(n)}(x_2|x_1) \cdot \tilde{\xi}_1(x_1) & \text{si } x_1 \neq x_1^{(n)} \\ \xi_{2|1}^{(n+1)}(x_2|x_1^{(n)}) \cdot \tilde{\xi}_1(x_1^{(n)}) & \text{si } x_1 = x_1^{(n)} \end{cases}$$

Por tanto;

$$\xi^{(n+1)}(x_1, x_2)$$

$$\begin{aligned} &= \xi_{2|1}^{(n)}(x_2|x_1) \cdot \tilde{\xi}_1(x_1) - \xi_{2|1}^{(n)}(x_2|x_1^{(n)}) \cdot \tilde{\xi}_1(x_1^{(n)}) \cdot \mathbf{1}_{x_1=x_1^{(n)}} \\ &\quad + \xi_{2|1}^{(n+1)}(x_2|x_1^{(n)}) \cdot \tilde{\xi}_1(x_1^{(n)}) \cdot \mathbf{1}_{x_1=x_1^{(n)}} \\ &= \xi^{(n)}(x_1, x_2) + \alpha_n \tilde{\xi}_1(x_1^{(n)}) \mathbf{1}_{x_1=x_1^{(n)}} [\xi_{(x_1^{(n)}, x_2^{(n)})} - \xi_{2|1}^{(n)}(x_2|x_1^{(n)})] \end{aligned}$$

donde

$$\mathbf{1}_{x_1=x_1^{(n)}} = \begin{cases} 1 & \text{si } x_1 = x_1^{(n)} \\ 0 & \text{si } x_1 \neq x_1^{(n)} \end{cases}$$

5. El algoritmo se detiene cuando

$$1 + \frac{\int_{\chi_1} \max_{x_2 \in \chi_2} [d(x_1, x_2, \xi^{(n)}) - m] \tilde{\xi}_1(x_1)}{\Phi[M(\xi^{(n)})]} \geq \delta$$

donde $0 < \delta < 1$ es una cota inferior para la eficiencia.

El uso de programas informáticos para la búsqueda de diseños óptimos supone un gran avance de modo que las aplicaciones prácticas aumentan.

Ejemplo 2.1.

Consideremos el modelo de regresión lineal

$$E[y(x)] = \gamma_1 x_1 + \gamma_2 x_2, \quad x_1 \in \chi_1 = \{0, 1\}, \quad x_2 \in \chi_2 = \{0, 1\}.$$

Aplicaremos el algoritmo anterior para mejorar el diseño inicial:

$$\xi^{(0)} = \left\{ \begin{array}{cccc} (0, 0) & (0, 1) & (1, 0) & (1, 1) \\ \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{2} \end{array} \right\}.$$

En la Tabla 2.1 se resumen las 10 primeras iteraciones.

Observamos que $\int_{\chi_1} \max_{x_2} d(x_1, x_2, \xi^{(n)}) d\tilde{\xi}_1(x_1)$ va tendiendo a 2, que es el número de parámetros.

Se alcanza una cota para la eficiencia del 98,6 % después de 30 iteraciones.

Diseño	$x^{(n)} = (x_1^{(n)}, x_2^{(n)})$	$\int_{\chi_1} \max_{x_2} d(x_1, x_2, \xi^{(n)}) d\tilde{\xi}_1(x_1) - m$	Efic
$\xi^{(0)} = \left\{ \begin{matrix} (0,0) & (0,1) & (1,0) & (1,1) \\ \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{2} \end{matrix} \right\}$	(1,0)	1.00	50 %
$\xi^{(1)} = \left\{ \begin{matrix} (0,0) & (0,1) & (1,0) & (1,1) \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{matrix} \right\}$	(0,1)	0.67	66.6 %
$\xi^{(2)} = \left\{ \begin{matrix} (0,0) & (0,1) & (1,0) & (1,1) \\ 0.17 & 0.33 & 0.25 & 0.25 \end{matrix} \right\}$	(0,1)	0.25	81.8 %
$\xi^{(3)} = \left\{ \begin{matrix} (0,0) & (0,1) & (1,0) & (1,1) \\ 0.1 & 0.4 & 0.25 & 0.25 \end{matrix} \right\}$	(0,1)	0.19	87.5 %
$\xi^{(4)} = \left\{ \begin{matrix} (0,0) & (0,1) & (1,0) & (1,1) \\ 0.08 & 0.42 & 0.25 & 0.25 \end{matrix} \right\}$	(0,1)	0.15	90.5 %
$\xi^{(5)} = \left\{ \begin{matrix} (0,0) & (0,1) & (1,0) & (1,1) \\ 0.07 & 0.43 & 0.25 & 0.25 \end{matrix} \right\}$	(0,1)	0.12	92.3 %
$\xi^{(6)} = \left\{ \begin{matrix} (0,0) & (0,1) & (1,0) & (1,1) \\ 0.06 & 0.44 & 0.25 & 0.25 \end{matrix} \right\}$	(0,1)	0.11	93.5 %
$\xi^{(7)} = \left\{ \begin{matrix} (0,0) & (0,1) & (1,0) & (1,1) \\ 0.05 & 0.45 & 0.25 & 0.25 \end{matrix} \right\}$	(0,1)	0.09	94.4 %
$\xi^{(8)} = \left\{ \begin{matrix} (0,0) & (0,1) & (1,0) & (1,1) \\ 0.05 & 0.45 & 0.25 & 0.25 \end{matrix} \right\}$	(0,1)	0.08	95.1 %
$\xi^{(9)} = \left\{ \begin{matrix} (0,0) & (0,1) & (1,0) & (1,1) \\ 0.04 & 0.45 & 0.25 & 0.25 \end{matrix} \right\}$	(0,1)	0.07	95.6 %
$\xi^{(10)} = \left\{ \begin{matrix} (0,0) & (0,1) & (1,0) & (1,1) \\ 0.04 & 0.46 & 0.25 & 0.25 \end{matrix} \right\}$	(0,1)	0.06	96.1 %

Tabla: 2.1: Diseños óptimos MR y sus eficiencias utilizando el algoritmo que encontramos en López Fidalgo y Garcet Rodríguez [Lóp04a].

Generalización de la cota de Atwood

Se han considerado anteriormente cotas para la D -eficiencia de dicho diseño MR basadas en el teorema de equivalencia. Dichas cotas pueden ser mejoradas para el criterio de D -optimización (Atwood [Atw69]) como vimos en el capítulo anterior.

En esta sección se obtienen cotas mejores en ese sentido para las D -eficiencias de diseños marginalmente restringidos (MR).

Teorema 2.8. *Si $M(\xi)$ es una matrix no singular, se verifica que:*

$$eff_{D-MR}(\xi) \geq \frac{m}{\int_{\chi_1} \max_{x_2} d(x_1, x_2, \xi) \eta_1(dx_1)}.$$

Demostración.

Como se hizo en el Teorema 1.10, puede suponerse que $M(\xi) = I$. De esta forma si η es un diseño D -óptimo MR, resulta que:

$$\begin{aligned} & \int_{\chi_1} \max_{x_2} d(x_1, x_2, \xi) \eta_1(dx_1) \\ & \geq \int_{\chi=\chi_1 \times \chi_2} d(x_1, x_2, \xi) \eta_{2|1}(dx_2|x_1) \eta_1(dx_1) \\ & = \int_{\chi} d(x_1, x_2, \xi) \eta(dx) = \int_{\chi} f^t(x_1, x_2) M^{-1}(\xi) f(x_1, x_2) \eta(dx) \\ & = \int_{\chi} Tr[f^t(x_1, x_2) M^{-1}(\xi) f(x_1, x_2)] \eta(dx) \\ & = Tr\left(M^{-1}(\xi) \int_{\chi} f^t(x_1, x_2) f(x_1, x_2) \eta(dx)\right) \\ & = Tr(I.M(\eta)) = Tr[M(\eta)]. \end{aligned}$$

Por tanto:

$$\begin{aligned} \left(\frac{|M(\xi)|}{|M(\eta)|}\right)^{1/m} &= \left(\frac{1}{|M(\eta)|}\right)^{1/m} \geq \frac{m}{Tr[M(\eta)]} \\ &\geq \frac{m}{\int_{\chi_1} \max_{x_2} d(x_1, x_2, \xi) \eta_1(dx_1)}, \end{aligned}$$

esto es,

$$ef_{D-MR}(\xi) \geq \frac{m}{\int_{\chi_1} \max_{x_2} d(x_1, x_2, \xi) \eta_1(dx_1)}.$$

□

La cota de la eficiencia del Teorema 2.8 es mayor que la cota proporcionada por López-Fidalgo y Garcet-Rodríguez [Lóp04a] para estos diseños:

$$2 - \frac{\int_{\chi_1} \max_{x_2} d(x_1, x_2, \xi) \eta_1(dx_1)}{m} \leq \frac{m}{\int_{\chi_1} \max_{x_2} d(x_1, x_2, \xi) \eta_1(dx_1)}$$

ya que

$$\begin{aligned} 2m \int_{\chi_1} \max_{x_2} d(x_1, x_2, \xi) \eta_1(dx_1) - \left(\int_{\chi_1} \max_{x_2} d(x_1, x_2, \xi) \eta_1(dx_1) \right)^2 &\leq m^2 \\ \Leftrightarrow 0 &\leq \left(m - \int_{\chi_1} \max_{x_2} d(x_1, x_2, \xi) \eta_1(dx_1) \right)^2. \end{aligned}$$

Ejemplo 2.2.

Consideremos el ejemplo anterior, $E[y(x)] = \gamma_1 x_1 + \gamma_2 x_2$ con $x_1 \in \chi_1 = \{0, 1\}$, $x_2 \in \chi_2 = \{0, 1\}$, con diseño marginal:

$$\tilde{\xi}_1 = \begin{Bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{Bmatrix}.$$

Para los diseños obtenidos en las dos primeras iteraciones:

$$\xi^{(1)} = \begin{Bmatrix} (0, 0) & (0, 1) & (1, 0) & (1, 1) \\ 1/4 & 1/4 & 0 & 1/2 \end{Bmatrix}$$

$$\xi^{(2)} = \begin{Bmatrix} (0, 0) & (0, 1) & (1, 0) & (1, 1) \\ 1/4 & 1/4 & 1/4 & 1/4 \end{Bmatrix},$$

puede observarse la mejoría para ambos diseños de la cota de Atwood y de la derivada del teorema de equivalencia.

Diseños	Antigua Cota	Nueva Cota
$\xi^{(1)}$	0.500	0.667
$\xi^{(2)}$	0.667	0.750

A-optimización

Sea la función Φ_A definida como

$$\Phi_A[M(\xi)] = Tr[M^{-1}(\xi)]. \quad (2.6)$$

Teniendo en cuenta que Φ_A es diferenciable sobre \mathcal{M}^+ , y que $\nabla[\Phi_A(M(\xi))] = [M^{-1}(\xi)]^2$, es fácil demostrar el siguiente:

Teorema 2.9. (Teorema de equivalencia) *Dada la función Φ_A , las siguientes condiciones son equivalentes:*

1. ξ^* es A-óptimo marginalmente restringido, $\xi^* \in \Xi_R^*$.

2.

$$\begin{aligned} & \int_{\chi_1} \max_{x_2 \in \chi_2} f^t(x_1, x_2) [M^{-1}(\xi^*)]^2 f(x_1, x_2) \tilde{\xi}_1(dx_1) = \\ & = \min_{\xi \in \Xi_R^+} \int_{\chi_1} \max_{x_2 \in \chi_2} f^t(x_1, x_2) [M^{-1}(\xi)]^2 f(x_1, x_2) \tilde{\xi}_1(dx_1). \end{aligned}$$

3.

$$\int_{\chi_1} \max_{x_2 \in \chi_2} f^t(x_1, x_2) [M^{-1}(\xi^*)]^2 f(x_1, x_2) \tilde{\xi}_1(dx_1) = Tr[M^{-1}(\xi^*)].$$

Donde $M_{(x_1, x_2)} = f(x_1, x_2)f^t(x_1, x_2)$ es la matriz de información del diseño $\xi \in \Xi$ concentrado en el punto $\mathbf{x} = (x_1, x_2)$.

Corolario 2.3. Sea ξ^* un diseño A -óptimo marginalmente restringido, entonces:

$$\int_{\chi_1} \max_{x_2 \in S_{\xi_2}} f^t(x_1, x_2)[M^{-1}(\xi^*)]^2 f(x_1, x_2) d\tilde{\xi}_1(x_1) = \text{Tr}[M^{-1}(\xi^*)].$$

Teorema 2.10. El diseño ξ^* es A -óptimo marginalmente restringido si y solo si para todo punto del soporte del diseño marginal $x_1 \in S_{\tilde{\xi}_1}$ se verifica que

$$\begin{aligned} & \max_{x \in \chi_2} f^t(x_1, x_2)[M^{-1}(\xi^*)]^2 f(x_1, x_2) \\ &= \int_{\chi_2} f^t(x_1, x_2)[M^{-1}(\xi^*)]^2 f(x_1, x_2) d\xi_{2|1}^*(dx_2|x_1). \end{aligned}$$

Teorema 2.11. Una condición suficiente para que un diseño $\xi^* \in \Xi_R$ sea óptimo simultáneamente respecto de los criterios D -y A -optimización es que:

$$k M^{-1}(\xi) = [M^{-1}(\xi)]^2. \quad (2.7)$$

Además en este caso:

$$\text{Tr}[M^{-1}(\xi^*)] = k \cdot m,$$

siendo k una cierta constante.

Demostración.

ξ^* es un diseño D -óptimo marginalmente restringido si y sólo si, aplicando el Teorema 2.6, minimiza

$$\int_{\chi_1} \max_{x_2 \in \chi_2} d(x_1, x_2, \xi) \tilde{\xi}_1(dx_1).$$

Entonces, por la condición 2.7, minimiza:

$$\int_{\chi_1} \max_{x_2 \in \chi_2} f^t(x_1, x_2) [M^{-1}(\xi)]^2 f(x_1, x_2) \tilde{\xi}_1(dx_1).$$

Por tanto, según el Teorema 2.9, ξ^* es A -óptimo marginalmente restringido.

Por otro lado:

$$\begin{aligned} Tr[M^{-1}(\xi^*)] &= \int_{\chi_1} \max_{x_2 \in \chi_2} f^t(x_1, x_2) [M^{-1}(\xi^*)]^2 f(x_1, x_2) \tilde{\xi}_1(dx_1) \\ &= k \int_{\chi_1} \max_{x_2 \in \chi_2} f^t(x_1, x_2) M^{-1}(\xi^*) f(x_1, x_2) \tilde{\xi}_1(dx_1) \\ &= k \cdot m \end{aligned}$$

como consecuencia del Teorema 2.6. \square

Algoritmo de construcción de diseños A -óptimos marginalmente restringidos

1. Establecemos un diseño inicial $\xi^{(0)} \in \Xi_R^+$, es decir, un diseño con matriz de información no singular.
Dado $\xi \in \Xi$, denotemos:

$$v(x_1, x_2, \xi) = f^t(x_1, x_2) [M^{-1}(\xi)]^2 f(x_1, x_2).$$

2. A partir de $\xi^{(n)}$ calculamos el punto $x^{(n)} = (x_1^{(n)}, x_2^{(n)})$ de $\chi_1 \times \chi_2$ (recuérdese que tomamos $\chi_1 = S_{\tilde{\xi}_1}$), siendo:
 $x^{(n)} =$

$$\arg \left\{ \max_{x_1 \in \chi_1 = S_{\tilde{\xi}_1}} \left\{ \max_{x_2 \in \chi_2} v(x_1, x_2, \xi^{(n)}) - \int_{\chi_2} v(x_1, x_2, \xi^{(n)}) \xi_{2|1}^{(n)}(dx_2|x_1) \right\} \right\}$$

3. Calculamos el diseño condicional en el nuevo punto obtenido:

$$\xi_{2|1}^{(n+1)}(x_2|x_1^{(n)}) = (1 - \alpha_n) \xi_{2|1}^{(n)}(x_2|x_1^{(n)}) + \alpha_n \mathbf{1}_{(x_1^{(n)}, x_2^{(n)})}$$

4. Finalmente, construimos el diseño conjunto:

$$\xi^{(n+1)}(x_1, x_2) = \begin{cases} \xi^{(n)}(x_1, x_2) = \xi_{2|1}^{(n)}(x_2|x_1) \cdot \tilde{\xi}_1(x_1) & \text{si } x_1 \neq x_1^{(n)} \\ \xi_{2|1}^{(n+1)}(x_2|x_1^{(n)}) \cdot \tilde{\xi}_1(x_1^{(n)}) & \text{si } x_1 = x_1^{(n)} \end{cases}$$

Por tanto

$$\begin{aligned} \xi^{(n+1)}(x_1, x_2) &= \xi_{2|1}^{(n)}(x_2|x_1) \cdot \tilde{\xi}_1(x_1) - \xi_{2|1}^{(n)}(x_2|x_1^{(n)}) \cdot \tilde{\xi}_1(x_1^{(n)}) \cdot \mathbf{1}_{x_1=x_1^{(n)}} \\ &\quad + \xi_{2|1}^{(n+1)}(x_2|x_1^{(n)}) \cdot \tilde{\xi}_1(x_1^{(n)}) \cdot \mathbf{1}_{x_1=x_1^{(n)}} \\ &= \xi^{(n)}(x_1, x_2) + \alpha_n \tilde{\xi}_1(x_1^{(n)}) \mathbf{1}_{x_1=x_1^{(n)}} (\xi_{(x_1^{(n)}, x_2^{(n)})} - \xi_{2|1}^{(n)}(x_2|x_1^{(n)})), \end{aligned}$$

siendo

$$\mathbf{1}_{x_1=x_1^{(n)}} = \begin{cases} 1 & \text{si } x_1 = x_1^{(n)} \\ 0 & \text{si } x_1 \neq x_1^{(n)} \end{cases}.$$

5. El algoritmo se detiene cuando

$$1 + \frac{\int_{\chi_1} \max_{x_2 \in \chi_2} \{v(x_1, x_2, \xi^{(n)}) - Tr[M^{-1}(\xi^{(n)})]\} \tilde{\xi}_1(x_1)}{\Phi[M(\xi^{(n)})]} \geq \delta,$$

donde $0 < \delta < 1$ es una cota inferior para la eficiencia.

2.3. Diseños Condicionalmente Restringidos

Como en la sección anterior consideremos diseños definidos sobre el espacio producto $\chi = \chi_1 \times \chi_2$, de modo que no todos los factores están bajo el control del experimentador. A diferencia de los diseños MR, en la práctica puede ocurrir que los valores de la variable no controlable, x_1 , sean desconocidos antes de que el experimento sea llevado a cabo. En este caso, supondremos que conocemos a priori una información sobre x_1 para diferentes valores específicos

de x_2 . Así, se puede establecer una distribución condicional $\tilde{\xi}_{1|2}$, y buscar diseños en $\chi = \chi_1 \times \chi_2$ con la restricción de que $\tilde{\xi}_{1|2}$ sea su distribución condicional, es decir, en este caso:

$$\Xi_R = \{\xi \in \Xi / \xi_{1|2}(\cdot | x_2) = \tilde{\xi}_{1|2}, x_2 \in S_{\xi_2}\}.$$

Usamos el soporte del diseño marginal en lugar del espacio de diseño porque la distribución condicional sólo tiene sentido para valores posibles de x_2 .

Se verifica que Ξ_R es convexo, es decir, si $\mu, \eta \in \Xi_R$ y $\xi = \alpha\mu + (1 - \alpha)\eta$ para $0 < \alpha < 1$ entonces $\xi_{1|2} = \alpha\mu_{1|2} + (1 - \alpha)\eta_{1|2} = \alpha\tilde{\xi}_{1|2} + (1 - \alpha)\tilde{\xi}_{1|2} = \tilde{\xi}_{1|2}$. De este modo, Ξ_R y \mathcal{M}_R son también conjuntos convexos.

En este caso, los generadores del conjunto \mathcal{M}_R , son las matrices de información de los diseños

$$\xi(x_1, x_2) = \begin{cases} \tilde{\xi}_{1|2}(x_1 | x_2^{(0)}) & \text{if } x_2 = x_2^{(0)} \\ 0 & \text{if } x_2 \neq x_2^{(0)} \end{cases}, \quad x_2^{(0)} \in \chi_2.$$

\mathcal{M}_R es un conjunto convexo en un espacio euclídeo de dimensión $\frac{m(m+1)}{2}$. El teorema de Caratheodory garantiza que, cualquier matriz de información es una combinación convexa de como mucho $1 + \frac{m(m+1)}{2}$ generadores. No existe garantía de encontrar un soporte finito para un diseño conjunto, pero el diseño ξ_2 tendrá como máximo $1 + \frac{m(m+1)}{2}$ puntos en su soporte. (Véase López Fidalgo y Garcet Rodríguez [Lóp04a]).

Lema 2.4. Sea $\xi \in \Xi_R^+$. Si Φ es diferenciable sobre \mathcal{M}^+ entonces

$$\begin{aligned} & \inf_{\eta \in \Xi_R} \partial\Phi[M(\xi), M(\eta)] \\ &= \inf_{x_2 \in \chi_2} \int_{\chi_1} \partial\Phi[M(\xi), M_{(x_1, x_2)}] d\tilde{\xi}_{1|2}(dx_1 | x_2), \end{aligned}$$

siendo $M_{(x_1, x_2)} = f(x_1, x_2)f^t(x_1, x_2)$ la matriz de información del diseño formado solamente por el punto (x_1, x_2) .

La demostración de este lema, la podemos encontrar en López Fidalgo y Garcet Rodríguez [Lóp04a].

Del lema anterior se sigue el teorema de equivalencia para diseños CR.

Teorema 2.12. *Si Φ es diferenciable en \mathcal{M}^+ las siguientes condiciones son equivalentes:*

1. $\Phi[M(\xi^*)] = \inf_{\xi \in \Xi_R} \Phi[M(\xi)].$
2.
$$\inf_{x_2 \in \chi_2} \int_{\chi_1} \partial \Phi[M(\xi^*), M_{(x_1, x_2)}] \tilde{\xi}_{1|2}(dx_1 \mid x_2) \\ = \sup_{\xi \in \Xi_R^+} \inf_{x_2 \in \chi_2} \int_{\chi_1} \partial \Phi[M(\xi), M_{(x_1, x_2)}] \tilde{\xi}_{1|2}(dx_1 \mid x_2).$$
3.
$$\inf_{x_2 \in \chi_2} \int_{\chi_1} \partial \Phi[M(\xi^*), M_{(x_1, x_2)}] \tilde{\xi}_{1|2}(dx_1 \mid x_2) = 0.$$

Demostración.

Considerando el Lema 2.4 y sustituyendo en el Teorema 2.4, obtenemos el resultado.

□

Cotas para la eficiencia.

Nuevamente, teniendo en cuenta que

$$eff_{\Phi-CR}(\xi) \geq 1 + \frac{\inf_{\eta \in \Xi_R} F_{\Phi}[M(\xi), M(\eta)]}{\Phi[M(\xi)]},$$

si la función criterio es diferenciable, por el Lema 2.4 resulta

$$\begin{aligned}
& ef f_{\Phi-CR}(\xi) \\
& \geq 1 + \frac{\inf_{\eta \in \Xi_R} F_{\Phi}[M(\xi), M(\eta)]}{\Phi[M(\xi)]} \\
& = 1 + \frac{\inf_{x_2 \in \chi_2} \int_{\chi_1} F_{\Phi}[M(\xi), f(x_1, x_2) f^t(x_1, x_2)] \tilde{\xi}_{1|2}(dx_1 | x_2)}{\Phi[M(\xi)]} \\
& = 1 + \frac{\inf_{x_2 \in \chi_2} \int_{\chi_1} \{f^t(x_1, x_2) \nabla \Phi[M(\xi)] f(x_1, x_2) - Tr[M(\xi) \nabla \Phi[M(\xi)]]\} \tilde{\xi}_{1|2}(dx_1 | x_2)}{\Phi[M(\xi)]}.
\end{aligned}$$

Algoritmos para Φ -óptimos CR

López-Fidalgo y Garcet-Rodríguez, [Lóp04a], proponen un algoritmo de tipo iterativo que proporciona una sucesión de diseños que, bajo ciertas condiciones, convergen al óptimo.

1. Se establece un diseño inicial $\xi^{(0)} \in \Xi_R^+$.
2. A partir de $\xi^{(n)}$ se calcula el punto $x_2^{(n)} \in \chi_2$, siendo este punto

$$x_2^{(n)} = \arg \left\{ \int_{\chi_1} \partial \Phi[M(\xi^{(n)}), M_{(x_1, x_2)}] \tilde{\xi}_{1|2}(dx_1 | x_2) \right\}.$$

3. Determinamos el nuevo diseño marginal:

$$\xi_2^{(n+1)}(x_2) = (1 - \alpha_n) \xi_2^{(n)}(x_2) + \alpha_n \mathbf{1}_{x_2^{(n)}}(x_2).$$

4. Finalmente, calculamos el diseño conjunto:

$$\xi^{(n+1)}(x_1, x_2) = \tilde{\xi}_{1|2}(x_1 | x_2) \xi_2^{(n+1)}(x_2).$$

5. El procedimiento se detiene cuando:

$$1 + \frac{\int_{\chi_1} \partial \Phi[M(\xi^{(n)}), M_{(x_1, x_2^{(n)})}] \xi_{1|2}^{(n)}(dx_1 | x_2^{(n)})}{\Phi[M(\xi^{(n)})]} \geq \delta,$$

donde $0 < \delta < 1$ es una cota inferior para la eficiencia.

Teorema 2.13. *Si Φ es una función criterio tal que*

1. Φ está acotada inferiormente y alcanza su mínimo en $M(\xi^*)$,
2. Φ tiene las derivadas de segundo orden acotadas, i.e., existe $B > 0$, tal que $|\partial^2 \Phi / \partial M_{ij} \partial M_{rt}| \leq B < \infty$,
3. $0 < \alpha_n < 1$, $\lim_{n \rightarrow \infty} \alpha_n = 0$, $\sum_{n=0}^{\infty} \alpha_n = \infty$ y $\sum_{n=0}^{\infty} \alpha_n^2 \leq k < \infty$,

entonces $\lim_{n \rightarrow \infty} \Phi[M(\xi^{(n)})] = \Phi[M(\xi^*)]$.

López-Fidalgo y Garcet-Rodríguez [Lóp04a] proporcionan una prueba de este teorema de convergencia.

2.3.1. Criterios de optimalidad clásicos para diseños CR

D –optimización.

Sea $\Phi_D[M(\xi)] = -\log|M(\xi)|$. Teniendo en cuenta que

$$F_{\Phi_D}[M(\xi), M_{(x_1, x_2)}] = -[d(x_1, x_2, \xi) - m],$$

y que Φ_D diferenciable en \mathcal{M}^+ , el teorema de equivalencia para diseños D –óptimos CR puede ser enunciado del siguiente modo:

Teorema 2.14. *Dada la función Φ_D las siguientes proposiciones son equivalentes:*

1. ξ^* es D –óptimo condicionalmente restringido, $\xi^* \in \Xi_R^*$.

$$\begin{aligned}
2. \quad & \sup_{x_2 \in \chi_2} \int_{\chi_1} d(x_1, x_2, \xi^*) \tilde{\xi}_{1|2}(dx_1|x_2) \\
& = \inf_{\xi \in \Xi_R^+} \sup_{x_2 \in \chi_2} \int_{\chi_1} d(x_1, x_2, \xi) \tilde{\xi}_{1|2}(dx_1|x_2),
\end{aligned}$$

siendo $d(x_1, x_2, \xi)$ la función de dispersión.

$$3. \quad \sup_{x_2 \in \chi_2} \int_{\chi_1} d(x_1, x_2, \xi^*) \tilde{\xi}_{1|2}(dx_1|x_2) = m.$$

Cota de la eficiencia

La cota para la eficiencia, viene dada por:

$$eff_{D-CR}(\xi) \geq 1 + \frac{\inf_G \text{generadores de } M_R F_{\Phi_D}[M(\xi), G]}{\Phi_D[M(\xi)]}.$$

Para garantizar la homogeneidad del criterio, se suele utilizar, para el cálculo de la eficiencia, la función

$$\Phi_D[M(\xi)] = |M(\xi)|^{-1/m}.$$

Por el teorema de equivalencia, para funciones diferenciables y por el Lema 2.4

$$eff_{D-CR}(\xi)$$

$$\begin{aligned}
& \geq 1 + \frac{\inf_{\eta \in \Xi_R} F_{\Phi}[M(\xi), M(\eta)]}{\Phi[M(\xi)]} \\
& = 1 + \frac{\inf_{x_2 \in \chi_2} \int_{\chi_1} F_{\Phi}[M(\xi), f(x_1, x_2) f^t(x_1, x_2)] \tilde{\xi}_{1|2}(dx_1|x_2)}{\Phi[M(\xi)]} \\
& = 1 + \frac{\inf_{x_2 \in \chi_2} \int_{\chi_1} [|M(\xi)|^{-1/m} \{ 1 - \frac{f^t(x_1, x_2) M^{-1}(\xi) f(x_1, x_2)}{m} \}] \tilde{\xi}_{1|2}(dx_1|x_2)}{|M(\xi)|^{-1/m}}
\end{aligned}$$

Operando, resulta que:

$$eff_{D-CR}(\xi) \geq 2 - \frac{1}{m} \max_{x_2 \in \chi_2} \int_{\chi_1} \tilde{\xi}_{1|2}(dx_1 | x_2) f^t(x_1, x_2) M^{-1}(\xi) f(x_1, x_2).$$

Algoritmo de construcción de diseños D-óptimos CR

1. Se establece un diseño inicial $\xi^{(0)} \in \Xi_R^+$.
2. A partir de $\xi^{(n)}$ se calcula el punto $x_2^{(n)} \in \chi_2$, siendo este el punto

$$x_2^{(n)} = \arg \max_{x_2 \in \chi_2} \int_{\chi_1} d(x_1, x_2, \xi^{(n)}) \tilde{\xi}_{1|2}(dx_1 | x_2).$$

3. Determinamos el nuevo diseño marginal:

$$\xi_2^{(n+1)}(x_2) = (1 - \alpha_n) \xi_2^{(n)}(x_2) + \alpha_n \mathbf{1}_{x_2^{(n)}}(x_2).$$

4. Finalmente, calculamos el diseño conjunto:

$$\xi^{(n+1)}(x_1, x_2) = \tilde{\xi}_{1|2}(x_1 | x_2) \xi_2^{(n+1)}(x_2).$$

5. El procedimiento se detiene cuando:

$$1 - \frac{\int_{\chi_1} [d(x_1, x_2^{(n)}, \xi^{(n)}) - m] \xi_{1|2}^{(n)}(dx_1 | x_2^{(n)})}{\Phi[M(\xi^{(n)})]} \geq \delta,$$

donde $0 < \delta < 1$.

Ejemplo 2.3.

Para aclarar los conceptos anteriores, consideremos el modelo de regresión:

$$E[y(x)] = \gamma_1 x_1 + \gamma_2 x_2, \quad x_1 \in \chi_1 = \{0, 1\}, \quad x_2 \in \chi_2 = \{0, 1\}.$$

La variable respuesta depende de dos variables explicativas, una controlable, x_2 , y otra no controlable y desconocida antes de que el experimento sea realizado, x_1 .

Supongamos que tenemos una información a priori sobre x_1 para diferentes valores de x_2 ,

$$\tilde{\xi}_{1|2}(\cdot|0) = \left\{ \begin{array}{cc} 0 & 1 \\ 1/4 & 3/4 \end{array} \right\}, \quad \tilde{\xi}_{1|2}(\cdot|1) = \left\{ \begin{array}{cc} 0 & 1 \\ 1/8 & 7/8 \end{array} \right\}.$$

En el caso de D-optimización, el objetivo será el diseño conjunto con valor máximo del determinante de la matriz de información, sujeto a una distribución condicional, $\tilde{\xi}_{1|2}$, es decir, el diseño conjunto será de la forma:

$$\xi = \left\{ \begin{array}{cccc} (0,0) & (1,0) & (0,1) & (1,1) \\ \frac{1-p}{4} & \frac{3(1-p)}{4} & \frac{p}{8} & \frac{7p}{8} \end{array} \right\},$$

siendo el diseño que será llevado en la práctica

$$\xi_2 = \left\{ \begin{array}{cc} 0 & 1 \\ 1-p & p \end{array} \right\}.$$

Un simple cálculo, demuestra que el D-óptimo CR se alcanza en $p^* = \frac{24}{41}$, con valor del determinante, $\frac{9}{41}$.

A continuación, aplicamos el algoritmo al ejemplo anterior, considerando como diseño inicial:

$$\xi^{(0)} = \left\{ \begin{array}{cccc} (0,0) & (1,0) & (0,1) & (1,1) \\ \frac{1}{8} & \frac{3}{8} & \frac{1}{16} & \frac{7}{16} \end{array} \right\},$$

es decir,

$$\xi_2^{(0)} = \left\{ \begin{array}{cc} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{array} \right\}.$$

La Tabla 2.2, proporciona un resumen de las 6 primeras iteraciones con los diseños y sus eficiencias

Podemos observar que el algoritmo no es monótono, y que la convergencia es lenta, probablemente porque partimos de un diseño cercano al óptimo.

Diseño conjunto	$x_2^{(n)}$	Eficiencia
$\xi^{(1)} = \left\{ \begin{array}{cccc} (0,0) & (0,1) & (1,0) & (1,1) \\ \frac{1}{8} & \frac{1}{16} & \frac{3}{8} & \frac{7}{16} \end{array} \right\}$	1	87.27 %
$\xi^{(2)} = \left\{ \begin{array}{cccc} (0,0) & (0,1) & (1,0) & (1,1) \\ \frac{1}{16} & \frac{3}{32} & \frac{3}{16} & \frac{21}{32} \end{array} \right\}$	0	60.86 %
$\xi^{(3)} = \left\{ \begin{array}{cccc} (0,0) & (0,1) & (1,0) & (1,1) \\ \frac{1}{8} & \frac{1}{16} & \frac{3}{8} & \frac{7}{16} \end{array} \right\}$	1	87.27 %
$\xi^{(4)} = \left\{ \begin{array}{cccc} (0,0) & (0,1) & (1,0) & (1,1) \\ \frac{3}{32} & \frac{5}{64} & \frac{9}{32} & \frac{35}{64} \end{array} \right\}$	0	92.73 %
$\xi^{(5)} = \left\{ \begin{array}{cccc} (0,0) & (0,1) & (1,0) & (1,1) \\ \frac{1}{8} & \frac{3}{8} & \frac{1}{16} & \frac{7}{16} \end{array} \right\}$	1	87.27 %
$\xi^{(6)} = \left\{ \begin{array}{cccc} (0,0) & (0,1) & (1,0) & (1,1) \\ \frac{5}{48} & \frac{7}{96} & \frac{5}{16} & \frac{49}{96} \end{array} \right\}$	1	99.75 %

Tabla: 2.2 : Diseños D-óptimos CR y sus eficiencias.

Generalización de la cota de Atwood

La cota para la eficiencia obtenida para los diseños D -óptimos CR, puede ser mejorada mediante una generalización de la cota propuesta por Atwood, [Atw69].

Teorema 2.15. *Si $M(\xi)$ es una matriz no singular, se cumple que*

$$eff_{D-CR}(\xi) \geq \frac{m}{\int_{\chi_1} \max_{x_2} d(x_1, x_2, \xi) \eta_{1|2}(dx_1|x_2)}.$$

Demostración.

Como se hizo en el Teorema 2.8, puede suponerse $M(\xi) = I$, y considerando un diseño D -óptimo CR, η :

$$\begin{aligned}
& \int_{\chi_1} \max_{x_2} d(x_1, x_2, \xi) \eta_{1|2}(dx_1|x_2) \\
& \geq \int_{\chi=\chi_1 \times \chi_2} d(x_1, x_2, \xi) \eta_{1|2}(dx_1|x_2) \eta_2(dx_2) \\
& = \int_{\chi} d(x_1, x_2, \xi) \eta(dx) = \int_{\chi} f^t(x_1, x_2) M^{-1}(\xi) f(x_1, x_2) \eta(dx) \\
& = \int_{\chi} \text{Tr}[f^t(x_1, x_2) M^{-1}(\xi) f(x_1, x_2)] \eta(dx) \\
& = \text{Tr} \left(M^{-1}(\xi) \int_{\chi} f^t(x_1, x_2) f(x_1, x_2) \eta(dx) \right) \\
& = \text{Tr}(I.M(\eta)) = \text{Tr}[M(\eta)].
\end{aligned}$$

Por tanto:

$$\begin{aligned}
\left(\frac{|M(\xi)|}{|M(\eta)|} \right)^{1/m} &= \left(\frac{1}{M(\eta)} \right)^{1/m} \geq \frac{m}{\text{Tr}[M(\eta)]} \\
&\geq \frac{m}{\int_{\chi_1} \max_{x_2} d(x_1, x_2, \xi) \eta_{1|2}(dx_1|x_2)},
\end{aligned}$$

esto es,

$$\text{eff}_{D-CR}(\xi) \geq \frac{m}{\int_{\chi_1} \max_{x_2} d(x_1, x_2, \xi) \eta_{1|2}(dx_1|x_2)}.$$

□

De nuevo, la cota de la eficiencia proporcionada en el teorema anterior es mayor que la cota dada por López-Fidalgo y Garcet-Rodríguez, [Lóp04a], para estos diseños. En efecto:

$$0 \leq \left(m - \int_{\chi_1} \max_{x_2} d(x_1, x_2, \xi) \eta_{1|x_2}(dx_1|x_2) \right)^2,$$

es decir:

$$2m \int_{\chi_1} \max_{x_2} d(x_1, x_2, \xi) \eta_{1|2}(dx_1 | x_2) - \left(\int_{\chi_1} \max_{x_2} d(x_1, x_2, \xi) \eta_{1|2}(dx_1 | x_2) \right)^2 \leq m^2.$$

Dividiendo por $m \int_{\chi_1} \max_{x_2} d(x_1, x_2, \xi) \eta_{1|2}(dx_1 | x_2)$, obtenemos:

$$2 - \frac{\int_{\chi_1} \max_{x_2} d(x_1, x_2, \xi) \eta_{1|2}(dx_1 | x_2)}{m} \leq \frac{m}{\int_{\chi_1} \max_{x_2} d(x_1, x_2, \xi) \eta_{1|2}(dx_1 | x_2)}.$$

Ejemplo 2.4.

Sea $E[y(x)] = \gamma_1 x_1 + \gamma_2 x_2$, donde $x_1 \in \chi_1 = \{0, 1\}$, $x_2 \in \chi_2 = \{0, 1\}$, con diseños condicionales:

$$\tilde{\xi}_{1|2}(\cdot | 0) = \left\{ \begin{array}{cc} 0 & 1 \\ 1/4 & 3/4 \end{array} \right\}, \text{ y } \tilde{\xi}_{1|2}(\cdot | 1) = \left\{ \begin{array}{cc} 0 & 1 \\ 1/8 & 7/8 \end{array} \right\}$$

y consideremos el diseño conjunto

$$\xi = \left\{ \begin{array}{cccc} (0, 0) & (0, 1) & (1, 0) & (1, 1) \\ 1/16 & 3/32 & 3/16 & 21/32 \end{array} \right\}.$$

En este caso la cota de la eficiencia proporcionada por López Fidalgo y Garcet Rodríguez [Lóp04a] es 0.77, y la nueva es de 0.82.

A-optimización

Consideremos la función criterio

$$\Phi_A[M(\xi)] = \text{Tr}[M^{-1}(\xi)].$$

En este caso, sabemos que Φ_A es diferenciable en \mathcal{M}^+ y

$$F_{\Phi_A}[M(\xi), M_{(x_1, x_2)}] = -[f^t(x_1, x_2)(M^{-1}(\xi))^2 f(x_1, x_2) - \text{Tr} M^{-1}(\xi)].$$

De modo que el teorema de equivalencia puede ser enunciado del siguiente modo:

Teorema 2.16. *Dada la función Φ_A , las siguientes proposiciones son equivalentes:*

1. ξ^* es A -óptimo condicionalmente restringido: $\xi^* \in \Xi_R^*$.
2.
$$\sup_{x_2 \in \chi_2} \int_{\chi_1} f^t(x_1, x_2) (M^{-1}(\xi^*))^{-2} f(x_1, x_2) \tilde{\xi}_{1|2}(dx_1 | x_2)$$
$$= \inf_{\xi \in \Xi_R^+} \sup_{x_2 \in \chi_2} \int_{\chi_1} f^t(x_1, x_2) (M^{-1}(\xi))^{-2} f(x_1, x_2) \tilde{\xi}_{1|2}(dx_1 | x_2).$$
3.
$$\sup_{x_2 \in \chi_2} \int_{\chi_1} f^t(x_1, x_2) (M^{-1}(\xi^*))^{-2} f(x_1, x_2) \tilde{\xi}_{1|2}(dx_1 | x_2)$$
$$= Tr[M^{-1}(\xi^*)].$$

Algoritmo de construcción de diseños A -óptimos CR

1. Se establece un diseño inicial $\xi^{(0)} \in \Xi_R^+$.
2. A partir de $\xi^{(n)}$ se calcula el punto $x_2^{(n)} \in \chi_2$, siendo este el punto

$$x_2^{(n)} = \arg \max_{x_2 \in \chi_2} \int_{\chi_1} f^t(x_1, x_2) (M^{-1}(\xi^{(n)}))^{-2} f(x_1, x_2) \tilde{\xi}_{1|2}(dx_1 | x_2).$$

3. Determinamos el nuevo diseño marginal:

$$\xi_2^{(n+1)}(x_2) = (1 - \alpha_n) \xi_2^{(n)}(x_2) + \alpha_n \mathbf{1}_{x_2^{(n)}}(x_2).$$

4. Finalmente, calculamos el diseño conjunto:

$$\xi^{(n+1)}(x_1, x_2) = \tilde{\xi}_{1|2}(x_1 | x_2) \xi_2^{(n+1)}(x_2).$$

5. El procedimiento se detiene cuando:

$$1 - \frac{\int_{\chi_1} [f^t(x_1, x_2^{(n)}) (M^{-1}(\xi^{(n)}))^{-2} f(x_1, x_2^{(n)}) - Tr(M^{-1}(\xi^{(n)}))] \tilde{\xi}_{1|2}^{(n)}(dx_1 | x_2^{(n)})}{\Phi[M(\xi^{(n)})]} \geq \delta,$$

donde $0 < \delta < 1$ es una cota inferior para la eficiencia.

2.4. Diseños Marginal y Condicionalmente Restringidos

Planteamos el caso en el que tenemos los dos tipos de restricciones anteriores. Ahora aparecen tres clases de variables independientes en el modelo:

- variables no controlables con valores conocidos antes de que el experimento sea realizado, $x_1 \in \chi_1$.
- variables no controlables con valores desconocidos antes de que el experimento sea realizado, $x_2 \in \chi_2$.
- variables controlables, $x_3 \in \chi_3$.

El espacio de diseño es el espacio producto $\chi = \chi_1 \times \chi_2 \times \chi_3$. Como restricciones tendremos un diseño marginal, $\tilde{\xi}_1$ y otro condicional $\tilde{\xi}_{2|13}$, de modo que el conjunto de diseños es

$$\Xi_R = \{\xi / \xi_1 = \tilde{\xi}_1, \xi_{2|13}(\cdot | x_1, x_3) = \tilde{\xi}_{2|13}(\cdot | x_1, x_3), (x_1, x_3) \in S_{\xi_{13}}\}.$$

Este conjunto es convexo, y la familia de matrices de información asociada, $\mathcal{M}_R = \{M(\xi) : \xi \in \Xi_R\}$, tiene las mismas propiedades que la familia de todas las matrices de información (Fedorov [Fed72]).

Usamos como soporte $S_{\xi_{13}}$ en vez de $\chi_1 \times \chi_3$ porque aunque $\tilde{\xi}_{2|13}(x_2 | x_1, x_3)$ existe para cualquier $x_1 \in \chi_1$ y $x_3 \in \chi_3$, el diseño que se lleva a cabo en la práctica $\xi_{3|1}$ podría anularse para algunos valores de x_3 .

Lema 2.5. Sea $\xi \in \Xi_R^+$. Si Φ es diferenciable en \mathcal{M}^+ entonces

$$\begin{aligned} & \inf_{\eta \in \Xi_R} \partial \Phi[M(\xi), M(\eta)] \\ &= \int_{\chi_1} \inf_{x_3 \in \chi_3} \int_{\chi_2} \partial \Phi[M(\xi), M_x] \tilde{\xi}_{2|13}(dx_2 | x_1, x_3) \tilde{\xi}_1(dx_1), \end{aligned}$$

siendo $M_x = f(x)f'(x)$ la matriz de información del diseño formado por el punto $x = (x_1, x_2, x_3)$.

Usando el lema anterior, se demuestra fácilmente el siguiente

Teorema 2.17. *Si Φ es diferenciable en \mathcal{M}^+ las siguientes condiciones son equivalentes:*

1. $\Phi[M(\xi^*)] = \inf_{\xi \in \Xi_R} \Phi[M(\xi)]$.
2.
$$\int_{\chi_1} \inf_{x_3 \in \chi_3} \int_{\chi_2} \partial \Phi[M(\xi^*), M_x] \tilde{\xi}_{2|13}(dx_2 | x_1, x_3) \tilde{\xi}_1(dx_1) \\ = \sup_{\xi \in \Xi_R^+} \int_{\chi_1} \inf_{x_3 \in \chi_3} \int_{\chi_2} \partial \Phi[M(\xi), M_x] \tilde{\xi}_{2|13}(dx_2 | x_1, x_3) \tilde{\xi}_1(dx_1).$$
3.
$$\int_{\chi_1} \inf_{x_3 \in \chi_3} \int_{\chi_2} \partial \Phi[M(\xi^*), M_x] \tilde{\xi}_{2|13}(dx_2 | x_1, x_3) \tilde{\xi}_1(dx_1) = 0.$$

Cota para la eficiencia.

Una vez más puesto que

$$eff_{\Phi-MCR}(\xi) \geq 1 + \frac{\inf_{\eta \in \Xi_R} F_{\Phi}[M(\xi), M(\eta)]}{\Phi[M(\xi)]},$$

si la función Φ es diferenciable, sabemos que la derivada direccional es lineal, por tanto:

$$\begin{aligned} & eff_{\Phi-MCR}(\xi) \\ & \geq 1 + \frac{\inf_{\eta \in \Xi_R} F_{\Phi}[M(\xi), M(\eta)]}{\Phi[M(\xi)]} \\ & = 1 + \frac{\int_{\chi_1} \inf_{x_3 \in \chi_3} \int_{\chi_2} F_{\Phi}[M(\xi), M_x] \tilde{\xi}_{2|13}(dx_2 | x_1, x_3) \tilde{\xi}_1(dx_1)}{\Phi[M(\xi)]} \\ & = 1 + \frac{\int_{\chi_1} \inf_{x_3 \in \chi_3} \int_{\chi_2} \{f^t(x_1, x_2, x_3) \nabla \Phi[M(\xi)] f(x_1, x_2, x_3) - Tr[M(\xi) \nabla \Phi[M(\xi)]]\}}{\Phi[M(\xi)]} \end{aligned}$$

$$\frac{\tilde{\xi}_{2|13}(dx_2 | x_1, x_3) \tilde{\xi}_1(dx_1)}{\Phi[M(\xi)]}.$$

Algoritmos para Φ –óptimos MCR

Un algoritmo iterativo para la construcción de diseños Φ –óptimos MCR es propuesto por López-Fidalgo y Garcet-Rodríguez, [Lóp04a], donde se demuestra también su convergencia.

1. Se establece un diseño inicial $\xi^{(0)} \in \Xi_R^+$.
2. A partir del diseño $\xi^{(n)} \in \Xi_R$ se calcula el punto $x_3^{(n)}(x_1)$ de χ_3 que es aquél donde el mínimo

$$\int_{\chi_2} \partial\Phi[M(\xi^{(n)}), M_x] \tilde{\xi}_{2|13}(dx_2 | x_1, x_3)$$

es alcanzado.

3. Se determina el nuevo diseño condicional,

$$\xi_{3|1}^{(n+1)}(x_3 | x_1) = (1 - \alpha_n) \xi_{3|1}^{(n)}(x_3 | x_1) + \alpha_n \mathbf{1}_{x_3^{(n)}(x_1)}(x_3).$$

4. Se define el nuevo diseño conjunto como

$$\xi^{(n+1)}(x) = \tilde{\xi}_1(x_1) \tilde{\xi}_{2|13}(x_2 | x_1, x_3) \xi_{3|1}^{(n+1)}(x_3 | x_1).$$

5. El procedimiento se detiene cuando

$$1 + \frac{\int_{\chi_1} \int_{\chi_2} \partial\Phi[M(\xi^{(n)}), M_{(x_1, x_2, x_3^{(n)})}] \tilde{\xi}_{2|13}(dx_2 | x_1, x_3^{(n)}) \tilde{\xi}_1(dx_1)}{\Phi[M(\xi^{(n)})]} \geq \delta,$$

donde $0 < \delta < 1$ es una cota inferior para la eficiencia.

2.4.1. Criterios de optimalidad clásicos para diseños MCR

D-optimización

Consideremos la función criterio $\Phi_D[M(\xi)] = -\log|M(\xi)|$. Sabemos que es diferenciable en \mathcal{M}_+ y que la derivada direccional es:

$$F_{\Phi_D}[M(\xi), M_{(x_1, x_2)}] = -[d(x_1, x_2, \xi) - m].$$

Teorema 2.18. *Dada la función Φ_D las siguientes proposiciones son equivalentes:*

1. ξ^* es D-óptimo marginal y condicionalmente restringido, $\xi^* \in \Xi_R^*$.
2.
$$\int_{\chi_1} \sup_{x_3 \in \chi_3} \int_{\chi_2} d(x_1, x_2, x_3, \xi^*) \tilde{\xi}_{2|13}(dx_2 | x_1, x_3) \tilde{\xi}_1(dx_1)$$

$$= \inf_{\xi \in \Xi_R^+} \int_{\chi_1} \sup_{x_3 \in \chi_3} \int_{\chi_2} d(x_1, x_2, x_3, \xi) \tilde{\xi}_{2|13}(dx_2 | x_1, x_3) \tilde{\xi}_1(dx_1).$$
3.
$$\int_{\chi_1} \sup_{x_3 \in \chi_3} \int_{\chi_2} d(x_1, x_2, x_3, \xi^*) \tilde{\xi}_{2|13}(dx_2 | x_1, x_3) \tilde{\xi}_1(dx_1) = m.$$

Cota para la eficiencia.

Para establecer una cota para la eficiencia, se suele usar como función criterio de D-optimización $\Phi_D[M(\xi)] = |M(\xi)|^{-1/m}$. En este caso, podemos escribir la cota como

$$\frac{\int_{\chi_1} \inf_{x_3 \in \chi_3} \int_{\chi_2} \{f^t(x_1, x_2, x_3) \nabla \Phi[M(\xi)] f(x_1, x_2, x_3) - \text{Tr}[M(\xi) \nabla \Phi[M(\xi)]]\}}{\frac{\tilde{\xi}_{2|13}(dx_2 | x_1, x_3) \tilde{\xi}_1(dx_1)}{\Phi[M(\xi)]}}$$

$$= 2 - \frac{1}{m} \int_{\chi_1} \max_{x_3} \int_{\chi_2} f^t(x_1, x_2, x_3) M^{-1}(\xi^{(n)}) f(x_1, x_2, x_3) \tilde{\xi}_{2|13}(dx_2 | x_1, x_3) \tilde{\xi}_1(dx_1).$$

Algoritmo de construcción de diseños D-óptimos MCR

1. Se establece un diseño inicial $\xi^{(0)} \in \Xi_R^+$.
2. A partir del diseño $\xi^{(n)} \in \Xi_R$ se calcula el punto $x_3^{(n)}(x_1)$ de χ_3 que es aquel donde el máximo

$$\int_{\chi_2} d(x_1, x_2, x_3, \xi^{(n)}) \tilde{\xi}_{2|13}(dx_2 | x_1, x_3)$$

es alcanzado.

3. Se determina el nuevo diseño condicional,

$$\xi_{3|1}^{(n+1)}(x_3 | x_1) = (1 - \alpha_n) \xi_{3|1}^{(n)}(x_3 | x_1) + \alpha_n \mathbf{1}_{x_3^{(n)}(x_1)}(x_3).$$

4. Se define el nuevo diseño conjunto como

$$\xi^{(n+1)}(x) = \tilde{\xi}_1(x_1) \tilde{\xi}_{2|13}(x_2 | x_1, x_3) \xi_{3|1}^{(n+1)}(x_3 | x_1).$$

5. El procedimiento se detiene cuando

$$1 - \frac{\int_{\chi_1} \int_{\chi_2} [d(x_1, x_2, x_3^{(n)}, \xi^{(n)}) - m] \tilde{\xi}_{2|13}(dx_2 | x_1, x_3^{(n)}) \tilde{\xi}_1(dx_1)}{\Phi[M(\xi^{(n)})]} \geq \delta,$$

donde $0 < \delta < 1$.

Presentamos un ejemplo para aclarar los conceptos expuestos anteriormente.

Ejemplo 2.5.

Consideremos el modelo de regresión lineal

$$E[y(x)] = \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 x_3,$$

donde $x_1 \in \chi_1 = \{0, 1\}$, $x_2 \in \chi_2 = \{0, 1\}$, $x_3 \in \chi_3 = \{0, 1\}$. La esperanza de la variable respuesta depende de tres variables explicativas, una variable controlable x_3 , otra no controlable con valores desconocidos antes de que el experimento sea realizado, x_2 y otra también desconocida pero cuyos valores son conocidos de antemano, x_1 , estos valores proporcionan el diseño marginal:

$$\tilde{\xi}_1 = \begin{Bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{Bmatrix}.$$

También suponemos que, tenemos un conocimiento a priori sobre la variable x_2 , así tenemos los diseños condicionales:

$$\tilde{\xi}_{2|13}(\cdot|0,0) = \left\{ \begin{array}{cc} 0 & 1 \\ 3/4 & 1/4 \end{array} \right\}, \quad \tilde{\xi}_{2|13}(\cdot|1,0) = \left\{ \begin{array}{cc} 0 & 1 \\ 2/3 & 1/3 \end{array} \right\},$$

y

$$\tilde{\xi}_{2|13}(\cdot|0,1) = \left\{ \begin{array}{cc} 0 & 1 \\ 4/5 & 1/5 \end{array} \right\}, \quad \tilde{\xi}_{2|13}(\cdot|1,1) = \left\{ \begin{array}{cc} 0 & 1 \\ 1/2 & 1/2 \end{array} \right\}.$$

En el caso de D-optimización, buscamos el diseño conjunto con mayor valor del determinante de la matriz de información, sujeto a la distribución marginal $\tilde{\xi}_1$ y a la condicional $\xi_{1|2}$, es decir, buscamos un diseño de la forma:

$$\xi = \left\{ \begin{array}{cccccccc} (0,0,0) & (1,0,0) & (0,1,0) & (0,0,1) & (1,1,0) & (1,0,1) & (0,1,1) & (1,1,1) \\ \frac{3(1-p)}{8} & \frac{1-q}{3} & \frac{1-p}{8} & \frac{2p}{5} & \frac{1-q}{6} & \frac{q}{4} & \frac{p}{10} & \frac{q}{4} \end{array} \right\},$$

donde

$$\xi_{3|1}(\cdot|0) = \left\{ \begin{array}{cc} 0 & 1 \\ 1-p & p \end{array} \right\} \quad \text{y} \quad \xi_{3|1}(\cdot|1) = \left\{ \begin{array}{cc} 0 & 1 \\ 1-q & q \end{array} \right\},$$

que son los diseños que se realizan en la práctica.

Aplicando el algoritmo al ejemplo dado obtenemos, después de 18 iteraciones, el diseño

$$\xi = \left\{ \begin{array}{cccccccc} (0,0,0) & (1,0,0) & (0,1,0) & (0,0,1) & (1,1,0) & (1,0,1) & (0,1,1) & (1,1,1) \\ 0.0104 & 0.1574 & 0.0035 & 0.03889 & 0.0787 & 0.1319 & 0.0972 & 0.1319 \end{array} \right\},$$

que alcanza una eficiencia del 99 %.

Generalización de la cota de Atwood.

Al igual que se hizo para los diseños MR y CR, veamos que la generalización de la cota de Atwood mejora la cota anteriormente expuesta.

Teorema 2.19. *Sea $M(\xi)$ una matriz no singular, se verifica que:*

$$eff_{D-MCR}(\xi) \geq \frac{m}{\int_{\chi_1} \max_{x_3} \int_{\chi_2} d(x_1, x_2, \xi) \eta_{2|13}(dx_2 | x_1, x_3) \eta_1(dx_1)}.$$

Demostración.

Podemos suponer $M(\xi) = I$, y consideremos η un diseño D -óptimo MCR (véase el teorema 2.8).

$$\begin{aligned} & \int_{\chi_1} \max_{x_3} \int_{\chi_2} d(x_1, x_2, \xi) \eta_{2|13}(dx_2 | x_1, x_3) \eta_1(dx_1) \\ & \geq \int_{\chi} d(x_1, x_2, x_3, \xi) \eta(dx) = \int_{\chi} f^t(x_1, x_2, x_3) M^{-1}(\xi) f(x_1, x_2, x_3) \eta(dx) = \\ & = \int_{\chi} Tr[f^t(x_1, x_2, x_3) M^{-1}(\xi) f(x_1, x_2, x_3)] \eta(dx) = \\ & = Tr\left(M^{-1}(\xi) \int_{\chi} f^t(x_1, x_2, x_3) f(x_1, x_2, x_3) \eta(dx)\right) \\ & = Tr(I.M(\eta)) = Tr[M(\eta)]. \end{aligned}$$

Por tanto:

$$\begin{aligned} \left(\frac{|M(\xi)|}{|M(\eta)|}\right)^{1/m} &= \left(\frac{1}{M(\eta)}\right)^{1/m} \geq \frac{m}{Tr[M(\eta)]} \\ &\geq \frac{m}{\int_{\chi_1} \max_{x_3} \int_{\chi_2} d(x_1, x_2, \xi) \eta_{2|13}(dx_2 | x_1, x_3) \eta_1(dx_1)}, \end{aligned}$$

esto es,

$$eff_{D-MCR}(\xi) \geq \frac{m}{\int_{\chi_1} \max_{x_3} \int_{\chi_2} d(x_1, x_2, \xi) \eta_{2|13}(dx_2 | x_1, x_3) \eta_1(dx_1)}.$$

□

La cota de la eficiencia del Teorema anterior es mayor que la cota proporcionada por López-Fidalgo y Garcet-Rodríguez, [Lóp04a], para estos diseños. En efecto,

$$2 - \frac{\int_{\chi_1} \max_{x_3} \int_{\chi_2} d(x_1, x_2, \xi) \eta_{2|13}(dx_2 | x_1, x_3) \eta_1(dx_1)}{m} \leq \frac{m}{\int_{\chi_1} \max_{x_3} \int_{\chi_2} d(x_1, x_2, \xi) \eta_{2|13}(dx_2 | x_1, x_3) \eta_1(dx_1)}$$

es decir,

$$\begin{aligned} & 2 m \frac{m}{\int_{\chi_1} \max_{x_3} \int_{\chi_2} d(x_1, x_2, \xi) \eta_{2|13}(dx_2 | x_1, x_3) \eta_1(dx_1)} \\ & - \left(\frac{m}{\int_{\chi_1} \max_{x_3} \int_{\chi_2} d(x_1, x_2, \xi) \eta_{2|13}(dx_2 | x_1, x_3) \eta_1(dx_1)} \right)^2 \leq m^2 \\ \Leftrightarrow & 0 \leq \left(m - \frac{m}{\int_{\chi_1} \max_{x_3} \int_{\chi_2} d(x_1, x_2, \xi) \eta_{2|13}(dx_2 | x_1, x_3) \eta_1(dx_1)} \right)^2. \end{aligned}$$

A-optimización

Sea $\Phi_A[M(\xi)] = \text{Tr}[M^{-1}(\xi)]$, y denotemos $\underline{x} = (x_1, x_2, x_3)$. Por ser Φ_A diferenciable en \mathcal{M}_R^+ , se verifica que

$$F_{\Phi_A}[M(\xi), M_{(x_1, x_2)}] = -[f^t(\underline{x})(M^{-1}(\xi))^2 f(\underline{x}) - \text{Tr} M^{-1}(\xi)].$$

Teniendo en cuenta lo anterior enunciamos el teorema de equivalencia:

Teorema 2.20. *Dada la función Φ_A las siguientes proposiciones son equivalentes:*

1. ξ^* es A-óptimo marginal y condicionalmente restringido, $\xi^* \in \Xi_R^*$.
2.
$$\int_{\chi_1} \sup_{x_3 \in \chi_3} \int_{\chi_2} f^t(\underline{x})(M^{-1}(\xi^*))^2 f(\underline{x}) \tilde{\xi}_{2|13}(dx_2 | x_1, x_3) \tilde{\xi}_1(dx_1)$$

$$= \inf_{\xi \in \Xi_R^+} \int_{\chi_1} \sup_{x_3 \in \chi_3} \int_{\chi_2} f^t(\underline{x})(M^{-1}(\xi))^2 f(\underline{x}) \tilde{\xi}_{2|13}(dx_2 | x_1, x_3) \tilde{\xi}_1(dx_1).$$
3.
$$\int_{\chi_1} \sup_{x_3 \in \chi_3} \int_{\chi_2} f^t(\underline{x})(M^{-1}(\xi^*))^2 f(\underline{x}) \tilde{\xi}_{2|13}(dx_2 | x_1, x_3) \tilde{\xi}_1(dx_1)$$

$$= Tr[M^{-1}(\xi^*)].$$

Algoritmo de construcción de diseños A-óptimos MCR

1. Se establece un diseño inicial $\xi^{(0)} \in \Xi_R^+$.
2. A partir del diseño $\xi^{(n)} \in \Xi_R$ se calcula el punto $x_3^{(n)}(x_1)$ de χ_3 que es aquel donde el máximo

$$\int_{\chi_2} f^t(\underline{x})(M^{-1}(\xi^{(n)}))^2 f(\underline{x}) \tilde{\xi}_{2|13}(dx_2 | x_1, x_3)$$

es alcanzado.

3. Se determina el nuevo diseño condicional,

$$\xi_{3|1}^{(n+1)}(x_3 | x_1) = (1 - \alpha_n) \xi_{3|1}^{(n)}(x_3 | x_1) + \alpha_n \mathbf{1}_{x_3^{(n)}(x_1)}(x_3).$$

4. Se define el nuevo diseño conjunto como

$$\xi^{(n+1)}(x) = \tilde{\xi}_1(x_1) \tilde{\xi}_{2|13}(x_2 | x_1, x_3) \xi_{3|1}^{(n+1)}(x_3 | x_1).$$

5. El procedimiento se detiene cuando

$$1 - \frac{\int_{\chi_1} \int_{\chi_2} [f^t(x_1, x_2, x_3^{(n)})(M^{-1}(\xi^{(n)}))^2 f(x_1, x_2, x_3^{(n)}) - Tr M^{-1}(\xi^{(n)})]}{1} \geq \delta,$$

$$\frac{\tilde{\xi}_{2|13}(dx_2 | x_1, x_3^{(n)}) \tilde{\xi}_1(dx_1)}{\Phi[M(\xi^{(n)})]} \geq \delta,$$

donde $0 < \delta < 1$ es una cota inferior para la eficiencia.

Capítulo 3

Algoritmos

*El **objetivo** de este capítulo consiste en el cálculo de diseños óptimos cuando algunos factores no están bajo el control del experimentador mediante el uso de técnicas numéricas. Los valores de estas variables pueden ser conocidos o desconocidos antes de que el experimento sea llevado a cabo. Se han considerado distintos criterios para el cálculo de dichos diseños.*

*La principal **contribución** es proporcionar una clase de algoritmos eficientes, que serán probados para encontrar diseños óptimos restringidos. El mayor problema que encontramos en la teoría proporcionada en el capítulo anterior es que los algoritmos sugeridos por los teoremas de equivalencia suponen un “gran esfuerzo” y lo que es peor una convergencia lenta. En este capítulo proponemos un algoritmo basado en los trabajos de Torsney y Mandal, [Tor01], y Mandal y Torsney, [Man00] y [Man06]. En ellos, encontramos un tipo de algoritmos multiplicativos para resolver un problema de programación matemática. Este tipo de algoritmos serán usados en nuestro contexto para mejorar la construcción de diseños óptimos MR y MCR.*

Se proporciona una equivalencia entre $D-$ y D_s- optimización para diseños MR y MCR, facilitando su cálculo.

3.1. Introducción

En el capítulo 1 hemos introducido el por qué de la necesidad de algoritmos especiales para el cálculo de diseños óptimos. Asimismo, en el capítulo anterior indicábamos que para casos muy particulares en los que el conjunto χ , el modelo lineal y la función criterio Φ sean muy simples, es posible proporcionar una solución analítica al problema de la construcción del diseño óptimo. Técnicas iterativas han sido desarrolladas (López-Fidalgo y Garcet-Rodríguez, [Lóp04a]) para proporcionar una serie de diseños que convergen al óptimo.

En este capítulo, se propone una clase de algoritmos multiplicativos para resolver este problema. Estos algoritmos utilizan un tipo de iteración propuesta por Torsney [Tor77]. Un mayor estudio sobre este tipo de algoritmos lo podemos encontrar en Torsney [Tor83], Torsney y Alahmadi [Tor92], Torsney y Mandal [Tor01] y Mandal y Torsney [Man00].

En la sección 3.2 se hace una introducción de este tipo de algoritmos. Se presenta el tipo de iteración y sus propiedades.

La sección 3.3 está dedicada a la adaptación de los algoritmos multiplicativos al problema del cálculo de diseños Φ -óptimos marginalmente restringidos. Se muestra la equivalencia entre diseños D - y D_s -óptimos.

En la sección 4.4 hacemos un estudio análogo al de la sección anterior pero para diseños marginal y condicionalmente restringidos.

La última sección presenta el cálculo de diseños MR y MCR para ejemplos reales.

3.2. Planteamiento del problema

Como en Mandal y Torsney [Man00], consideremos el siguiente problema:

Minimizar la función criterio $\Phi(p)$ sujeto al conjunto de restricciones

$$\mathcal{P} \equiv \left\{ p = (p_1, \dots, p_J) : p_j \geq 0, \sum_{j=1}^J p_j = 1 \right\}. \quad (3.1)$$

Una iteración, propuesta por Torsney [Tor77], que conserva las restricciones del problema anterior, es la siguiente:

$$p_j^{(r+1)} = \frac{p_j^{(r)} f(d_j^{(r)})}{\sum_{i=1}^J p_i^{(r)} f(d_i^{(r)})}, \quad (3.2)$$

donde

$$d_j^{(r)} = \left. \frac{\partial \Phi}{\partial p_i} \right|_{p_i = p_j^{(r)}},$$

y $f(\cdot)$ satisface las siguientes condiciones:

1. Es positiva.
2. Es estrictamente creciente.
3. Puede depender de uno o varios parámetros.

Las iteraciones (3.2) poseen las siguientes propiedades (Torsney [Tor88]):

1. $p^{(r)}$ es factible.
2. $F_{\Phi}\{p^{(r)}, p^{(r+1)}\} \geq 0$. La igualdad se verifica cuando las d_j que corresponden a los p_j distintos de cero, tienen el mismo valor, d , es decir, $d_j = \sum p_i d_i = d$, y por tanto

$$p_j^{(r+1)} = \frac{p_j^{(r)} f(d_j)}{\sum_{i=1}^J p_i^{(r)} f(d_i)} = \frac{p_j^{(r)} f(d)}{f(d) \sum_{i=1}^J p_i^{(r)}} = p_j^{(r)}.$$

3. El soporte de $p^{(r+1)}$ está incluido en el soporte de $p^{(r)}$.

Otras propiedades interesantes pueden ser consultadas en Silvey, Titterington y Torsney [Sil78] y en Torsney y Mandal [Tor01].

Fedorov [Fed72] y Wynn [Wyn72] proponen un tipo de algoritmos que modifican un p_j y cambian los otros proporcionalmente. Estos algoritmos son útiles cuando muchos de los p_j son cero en el óptimo, como sucede en problemas de diseño. Por otro lado, cuando todos los p_j son positivos en el óptimo o cuando se sabe cuáles son positivos, iteraciones restringidas del tipo steepest ascent o Newton pueden ser apropiadas (véase Wu [Wu78] y Atwood [Atw76] y [Atw80]). Torsney [Tor83], sugirió que la iteración (3.2) podría ser útil cuando solamente unos pocos de los pesos del óptimo pudieran ser cero.

Anteriormente, mencionamos que $f(\cdot)$ puede depender de uno o más parámetros. Torsney y Alahmadi [Tor92] estudian estas ideas. Titterington [Tit76] demuestra la monotonicidad para $f(d) = d$ para D -optimización. Torsney [Tor83] estudia monotonicidad de valores particulares de δ . Torsney [Tor83] también establece una condición suficiente para la monotonicidad para $f(d) = d^\delta$, $\delta = \frac{1}{t+1}$, cuando la función criterio $\Phi(\xi)$ es homogénea de grado $-t$, $t > 0$ con derivadas positivas y prueba esta condición en el caso de criterios lineales tales como c - optimización y A - optimización cuando $t = 1$, es decir, $\delta = 1/2$. Cuando $\delta = 1$, se obtiene un algoritmo EM, cuya convergencia es lenta (véase Dempster [Dem77]).

3.3. Construcción de diseños óptimos MR

Consideremos la función de regresión $E[y(x)] = \eta^t(x_1, x_2)\theta$, $(x_1, x_2) \in \chi = \chi_1 \times \chi_2$, siendo $\eta^t(x_1, x_2) = (\eta_1(x_1, x_2), \dots, \eta_m(x_1, x_2))$ un vector de m funciones continuas linealmente independientes sobre χ ,

y θ el vector m -dimensional de parámetros desconocidos. Como se definió en la sección anterior, ξ^* es un diseño MR Φ -óptimo si

$$\min_{\xi \in \Xi_R} \Phi(\xi) = \Phi(\xi^*).$$

Supongamos que $\chi_1 = \{x_1^{(1)}, \dots, x_1^{(I)}\}$ y $\chi_2 = \{x_2^{(1)}, \dots, x_2^{(J)}\}$. El diseño aproximado ξ lo podemos escribir como

$$\xi(x_1^{(i)}, x_2^{(j)}) = \tilde{\xi}_1(x_1^{(i)}) \times \xi_{2|1}(x_2^{(j)} | x_1^{(i)}) \equiv \tilde{\xi}_{1i} \times \xi_{2j|1i}.$$

Por tanto, ξ esta caracterizado por el conjunto de pesos marginales,

$$\tilde{\xi}_1 = \left(\tilde{\xi}_1(x_1^{(1)}), \dots, \tilde{\xi}_1(x_1^{(I)}) \right) = (\tilde{\xi}_{11}, \dots, \tilde{\xi}_{1I}),$$

y para cada valor $x_1^{(i)}$, el objetivo será optimizar el conjunto de diseños condicionales

$$\xi_{2|1i} = \left(\xi_{2|1}(x_2^{(1)} | x_1^{(i)}), \dots, \xi_{2|1}(x_2^{(J)} | x_1^{(i)}) \right) = (\xi_{21|1i}, \dots, \xi_{2J|1i}).$$

Estos pesos, deben satisfacer las condiciones:

$$\begin{aligned} \tilde{\xi}_{1i} &\geq 0, \quad i = 1, \dots, I \\ \xi_{2j|1i} &\geq 0, \quad j = 1, \dots, J \text{ para cada } i = 1, \dots, I \end{aligned}$$

y

$$\sum_{i=1}^I \tilde{\xi}_{1i} = \sum_{j=1}^J \xi_{2j|1i} = 1, \quad i = 1, \dots, I.$$

Para cada $x_1^{(i)}$ seleccionamos $\xi_{2|1i}$ de manera óptima. Si $\hat{\theta}$ es el estimador mínimo cuadrático de θ , entonces $\Sigma_{\hat{\theta}} \propto M^{-1}(\xi) = M^{-1}(\tilde{\xi}_1 \xi_{2|1})$, siendo

$$M(\xi) = \sum_{i=1}^I \sum_{j=1}^J \tilde{\xi}_{1i} \xi_{2j|1i} \eta(x_1^{(i)}, x_2^{(j)}) \eta^t(x_1^{(i)}, x_2^{(j)})$$

la matriz de información.

Podemos generalizar el problema (3.1) para diseños marginalmente restringidos del siguiente modo:

$$\text{Minimizar } \Phi(\underline{p}_1, \dots, \underline{p}_I) \text{ sujeto a} \quad (3.3)$$

$$\mathcal{P} \equiv \left\{ p = (\underline{p}_1, \dots, \underline{p}_I) : \underline{p}_i = (p_{i1}, \dots, p_{iJ}) : p_{ij} \geq 0, \right. \\ \left. \sum_j p_{ij} = 1 \text{ para cada } i = 1, \dots, I \right\}.$$

Este problema tiene un conjunto de restricciones sobre p_{i1}, \dots, p_{iJ} ; estas variables deben ser no negativas y sumar uno para cada $i = 1, \dots, I$. En este contexto, estas variables p_{ij} son los pesos de los diseños condicionales $\xi_{2j|1i}$. De modo que la iteración (3.2) resulta ahora:

$$\xi_{2j|1i}^{(r+1)} = \frac{\xi_{2j|1i}^{(r)} f(d_{2j|1i}^{(r)})}{\sum_{t=1}^J \xi_{2t|1i}^{(r)} f(d_{2t|1i}^{(r)})}, \quad (3.4)$$

donde $d_{2j|1i} = \partial\Phi/\partial\xi_{2j|1i}$ y $f(\cdot)$ es positiva y estrictamente creciente.

Sea $F_{2j|1i} = d_{2j|1i} - \sum_{t=1}^J \xi_{2t|1i} d_{2t|1i}$. Las condiciones para un mínimo local son:

$$F_{2j|1i}^* \begin{cases} = 0, & \text{para } \xi_{2j|1i}^* > 0 \\ \geq 0, & \text{para } \xi_{2j|1i}^* = 0. \end{cases} \quad (3.5)$$

3.3.1. Construcción de diseños MR para distintas funciones criterio

En este apartado calculamos los pasos del algoritmo para las funciones criterio más utilizadas.

El cálculo de las derivadas $d_{2j|1i}$, para las diferentes funciones criterio, lo podemos encontrar en el apéndice A.1.

1. Diseños D -óptimos.

Consideremos la función, $\Phi(\xi) = \Psi[M(\xi)] = -\log \{\det\{M(\xi)\}\}$.

Las derivadas son

$$d_{2j|1i} = \frac{\partial \Phi}{\partial \xi_{2j|1i}} = -\tilde{\xi}_{1i} \eta^t(x_1^{(i)}, x_2^{(j)}) M^{-1}(\xi) \eta(x_1^{(i)}, x_2^{(j)}).$$

Usamos el algoritmo (3.4) y elegimos $f(d) = d$ pues con esta elección el algoritmo es monótono para determinar un diseño D -óptimo no restringido (Titterington [Tit76]). De este modo, esperamos que el algoritmo para el caso restringido, sea monótono para encontrar los diseños condicionales $\xi_{2j|1i}$.

2. Diseños c -óptimos.

Los diseños c -óptimos, como ya indicábamos en el capítulo 1, son adecuados cuando queremos estimar una combinación lineal de los parámetros $c^t \theta$ con mínima varianza. La función criterio que va a ser minimizada es $\Phi(\xi) = \Psi[M(\xi)] = c^t M^{-1}(\xi) c$, cuyas derivadas son:

$$d_{2j|1i} = \frac{\partial \Phi}{\partial \xi_{2j|1i}} = -\tilde{\xi}_{1i} [\eta^t(x_1^{(i)}, x_2^{(j)}) M^{-1}(\xi) c]^2.$$

En este caso, obtenemos los diseños condicionales usando el algoritmo (3.4) con $f(d) = d^{1/2}$,

$$\xi_{2j|1i}^{(r+1)} = \frac{\xi_{2j|1i}^{(r)} (d_{2j|1i}^{(r)})^{1/2}}{\sum_{t=1}^J \xi_{2t|1i}^{(r)} (d_{2t|1i}^{(r)})^{1/2}}.$$

La elección de esta función se debe nuevamente a que para c -óptimos no restringidos el algoritmo es monótono (Torsney [Tor83], y Fellman [Fel74]). Resultados experimentales muestran que también será monótona para encontrar diseños óptimos condicionales.

3. Diseños D_A -óptimos

Los diseños D_A -óptimos se calculan cuando el interés no está en todos los parámetros, sino solamente en k combinaciones

lineales de θ , por ejemplo $A\theta$, siendo $A \in \mathcal{M}_{k \times m}$ una matriz de rango $k < m$. La función criterio es

$$\Phi(\xi) = \Psi[M(\xi)] = \log \det \{AM^{-1}(\xi)A^t\}.$$

Las derivadas son

$$\begin{aligned} d_{2j|1i} &= \frac{\partial \Phi}{\partial \xi_{2j|1i}} \\ &= -\tilde{\xi}_{1i} \eta^t(x_1^{(i)}, x_2^{(j)}) M^{-1}(\xi) A^t [AM^{-1}(\xi)M^t]^{-1} AM^{-1}(\xi) \eta(x_1^{(i)}, x_2^{(j)}). \end{aligned}$$

Una vez mas, la experiencia sugiere considerar la función $f(d) = d$ en el algoritmo (3.4) para que sea monótono.

El resultado siguiente revela otra elección. Podemos reemplazar las derivadas de esta función criterio por las del D -óptimo.

Teorema 3.1. *Si $E[y(\underline{x})] = \underline{\theta}_1^t \underline{x}_1 + \underline{\theta}_2^t \underline{x}_2$, entonces el diseño D -óptimo marginalmente restringido es equivalente al diseño D_{θ_2} -óptimo.*

Demostración

Sea $M(\xi)$ la matriz de información. Podemos escribir

$$\begin{aligned} M(\xi) &= E \left[\begin{pmatrix} \underline{x}_1 \\ \underline{x}_2 \end{pmatrix} \begin{pmatrix} \underline{x}_1 & \underline{x}_2 \end{pmatrix} \right] \\ &= \begin{pmatrix} E[\underline{x}_1 \underline{x}_1^t] & E[\underline{x}_1 \underline{x}_2^t] \\ E[\underline{x}_2 \underline{x}_1^t] & E[\underline{x}_2 \underline{x}_2^t] \end{pmatrix} = \begin{pmatrix} M_{11} & M_{12} \\ M_{12}^t & M_{22} \end{pmatrix} \end{aligned}$$

Si $\hat{\underline{\theta}}$ es el estimador mínimo cuadrático de $\underline{\theta}$, entonces

$$\Sigma_{\hat{\underline{\theta}}} \propto M^{-1}(\xi) = \begin{pmatrix} N_{11} & N_{12} \\ N_{12}^t & N_{22} \end{pmatrix}$$

y

$$\Sigma_{\hat{\underline{\theta}_2}} \propto [M_{22} - M_{12}^t M_{11}^{-1} M_{12}]^{-1} = N_{22}.$$

De modo que

$$\det \{ \Sigma_{\hat{\underline{\theta}}} \} \propto \frac{1}{\det \{ M(\xi) \}} = \frac{\det \{ [M_{22} - M_{12}^t M_{11}^{-1} M_{12}]^{-1} \}}{\det \{ M_{11} \}}$$

$$= \frac{\det\{N_{22}\}}{\det\{M_{11}\}} \propto \frac{\det\{\Sigma_{\hat{\theta}_2}\}}{\det\{M_{11}\}}$$

Debido a que el diseño marginal $\tilde{\xi}_1$ es conocido, la matriz M_{11} puede ser considerada como una matriz constante, de modo que el mismo diseño maximiza $\det \Sigma$ y $\det \Sigma_2$.

□

4. Diseños L_A -óptimos

Podemos escribir la función criterio como

$$\Phi(\xi) = \Psi[M(\xi)] = \text{tr}\{AM^{-1}(\xi)A^t\}$$

siendo $A \in \mathcal{M}_{k \times m}$ una matriz de rango $k < m$. En este caso, las derivadas son

$$d_{2j|1i} = \frac{\partial \Phi}{\partial \xi_{2j|1i}} = -\tilde{\xi}_{1i} \eta^t(x_1^{(i)}, x_2^{(j)}) M^{-1}(\xi) A A^t M^{-1}(\xi) \eta(x_1^{(i)}, x_2^{(j)}).$$

Es de esperar que el algoritmo (3.4) con $f(d) = d^{1/2}$ sea monótono.

En todos los ejemplos de la siguiente sección consideramos, para iniciar el algoritmo, los diseños condicionales $\xi_{2j|1i}$ uniformes. Con la expresión “usamos el algoritmo 3.4” queremos indicar que se considera la versión apropiada del algoritmo anteriormente indicado.

3.3.2. Ejemplos

Ejemplo 1

Consideremos el modelo lineal $E[y(x)] = \theta_1 x_1 + \theta_2 x_2$, $\chi_1 = \{0, 1, 2\}$, $\chi_2 = \{0, 1\}$, y supongamos que el diseño marginal sigue una distribución binomial $\tilde{\xi}_1 \equiv Bi(2, 1/5)$. Como esperábamos, para $A^t = c^t = (0, 1)$, el algoritmo (3.4) converge de manera monótona a las

siguientes distribuciones para los criterios $D-$, $c-$, L_A- y D_A- óptimos:

$$\xi_{2|1}^*(\cdot | 0) = \begin{Bmatrix} 0 & 1 \\ 0 & 1 \end{Bmatrix}, \quad \xi_{2|1}^*(\cdot | 1) = \begin{Bmatrix} 0 & 1 \\ \frac{1}{4} & \frac{3}{4} \end{Bmatrix}$$

y

$$\xi_{2|1}^*(\cdot | 2) = \begin{Bmatrix} 0 & 1 \\ 1 & 0 \end{Bmatrix}.$$

Este ejemplo es una ilustración del resultado del Teorema 3.1. En este caso c -optimalidad y D_{θ_2} -optimalidad coinciden.

Aunque la monotonidad no garantiza una convergencia rápida, una buena elección de la función $f(d)$ proporciona diseños óptimos en pocas iteraciones. La Tabla 3.1 muestra que para $D-$, D_A- , $c-$ y L_A- optimización se consigue $F_{2j|1i} \geq -0.1$ en una o dos iteraciones.

Criterio	n=1	n=2	n=3	n=4
D -optimización	1	12	36	57
D_A -optimización	1	3	3	4
c -optimización	2	4	7	11
L_A -optimización	2	4	7	11

Tabla: 3.1 : Número de iteraciones necesarias para conseguir $\min_{1 \leq i \leq I, 1 \leq j \leq J} \{F_{2j|1i}\} \geq -10^{-n}$

En el caso continuo, $\chi_2 = [0, 1]$, consideramos el espacio de diseño aproximado por una red de 11 puntos equiespaciados en intervalos de amplitud 0.1 entre 0 y 1. Usando el algoritmo (3.4), obtenemos los mismos diseños condicionales que anteriormente.

Ejemplo 2

Consideremos ahora el modelo de regresión lineal con término independiente $E[y(x)] = \theta_0 + \theta_1 x_1 + \theta_2 x_2$, con $x_1 \in \chi_1 = \{0, 1, 2\}$, $x_2 \in \chi_2 = \{0, 1\}$ y $\xi_1 \equiv Bi(2, 1/5)$. Usando nuevamente el algoritmo (3.4), obtenemos los siguientes diseños condicionales:

$$\xi_{2|1}^*(\cdot | 0) = \begin{Bmatrix} 0 & 1 \\ 0.529 & 0.471 \end{Bmatrix}, \quad \xi_{2|1}^*(\cdot | 1) = \begin{Bmatrix} 0 & 1 \\ 0.384 & 0.616 \end{Bmatrix}$$

y

$$\xi_{2|1}^*(\cdot | 2) = \begin{Bmatrix} 0 & 1 \\ 0.964 & 0.036 \end{Bmatrix}.$$

En ambos ejemplos se ha comprobado que los diseños obtenidos son óptimos.

3.4. Construcción de diseños óptimos MCR

Como vimos en el capítulo anterior, en el modelo pueden aparecer tres clases de variables independientes. Sea $E[y(x)] = \eta^t(x_1, x_2, x_3)\theta$, donde $x_1 \in \chi_1$, $x_2 \in \chi_2$, $x_3 \in \chi_3$ y $\eta^t(x_1, x_2, x_3) = (\eta_1(x_1, x_2, x_3), \dots, \eta_m(x_1, x_2, x_3))$ con varianza constante. Estas variables pueden ser variables controlables, x_3 , no controlables con valores conocidos antes de que el experimento sea llevado a cabo, x_1 , y no controlables con valores desconocidos, x_2 .

Como ya definíamos en el capítulo anterior, un diseño ξ^* diremos que es Φ -óptimo si

$$\min_{\xi \in \Xi_R} \Phi(\xi) = \Phi(\xi^*).$$

Como en la sección anterior, sea $\chi = \chi_1 \times \chi_2 \times \chi_3$ con $\chi_1 = \{x_1^{(1)}, \dots, x_1^{(I)}\}$, $\chi_2 = \{x_2^{(1)}, \dots, x_2^{(J)}\}$ y $\chi_3 = \{x_3^{(1)}, \dots, x_3^{(K)}\}$. El diseño aproximado, ξ , lo podemos escribir como

$$\begin{aligned} \xi(x_1^{(i)}, x_2^{(j)}, x_3^{(k)}) &= \tilde{\xi}_1(x_1^{(i)}) \times \tilde{\xi}_{2|1,3}(x_2^{(j)} | x_1^{(i)}, x_3^{(k)}) \times \xi_{3|1}(x_3^{(k)} | x_1^{(i)}) \\ &= \tilde{\xi}_{1i} \times \tilde{\xi}_{2j|1i,3k} \times \xi_{3k|1i}. \end{aligned}$$

El diseño, ξ , queda caracterizado por los diseños marginales:

$$\tilde{\xi}_1 = (\tilde{\xi}_1(x_1^{(1)}), \dots, \tilde{\xi}_1(x_1^{(I)})) = (\tilde{\xi}_{11}, \dots, \tilde{\xi}_{1I}),$$

así, para cada $x_1^{(i)}, x_3^{(k)}$, los pesos condicionales son:

$$\begin{aligned}\tilde{\xi}_{2|1i,3k} &= (\tilde{\xi}_{2|13}(x_2^{(1)} | x_1^{(i)}, x_3^{(k)}), \dots, \tilde{\xi}_{2|13}(x_2^{(J)} | x_1^{(i)}, x_3^{(k)})) \\ &= (\tilde{\xi}_{21|1i,3k}, \dots, \tilde{\xi}_{2J|1i,3k}),\end{aligned}$$

y finalmente, los diseños condicionales para cada $x_1^{(i)}$:

$$\xi_{3|1i} = (\xi_{3|1}(x_3^{(1)} | x_1^{(i)}), \dots, \xi_{3|1}(x_3^{(K)} | x_1^{(i)})) = (\xi_{31|1i}, \dots, \xi_{3K|1i}).$$

Estos diseños satisfacen las siguientes condiciones:

$$\xi_{3k|1i} \geq 0, \quad k = 1, \dots, K \text{ para cada } i = 1, \dots, I,$$

y

$$\sum_{k=1}^K \xi_{3k|1i} = 1, \quad i = 1, \dots, I.$$

Como dijimos en el capítulo anterior, los diseños $\xi_{3|1i}$ son los que se realizan en la práctica, por esto, nuestro interés se centra en elegir los diseños $\xi_{3|1i}$ de manera óptima para cada $x_1^{(i)}$. La matriz de información de los diseños se puede escribir del siguiente modo

$$M(\xi) = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \tilde{\xi}_{1i} \tilde{\xi}_{2j|1i,3k} \xi_{3k|1i} \eta(x_1, x_2, x_3) \eta^t(x_1, x_2, x_3).$$

Una vez más, tenemos un ejemplo del problema (3.3), donde las variables p_{ik} se corresponden con los pesos de los diseños condicionales restringidos $\xi_{3k|1i}$. De nuevo, podemos utilizar el algoritmo multiplicativo. Adaptándolo a nuestro caso, resulta que la iteración es

$$\xi_{3k|1i}^{(r+1)} = \frac{\xi_{3k|1i}^{(r)} f(d_{3k|1i}^{(r)})}{\sum_{t=1}^J \xi_{3t|1i}^{(r)} f(d_{3t|1i}^{(r)})}, \quad (3.6)$$

donde $d_{3k|1i} = \partial \Phi / \partial \xi_{3k|1i}$ y $f(\cdot)$ es positiva y estrictamente creciente.

Sea $F_{3k|1i} = d_{3k|1i} - \sum_{t=1}^J \xi_{3t|1i} d_{3t|1i}$. Las condiciones de primer orden para el mínimo local son:

$$F_{3k|1i}^* \begin{cases} = 0, & \text{para } \xi_{3k|1i}^* > 0 \\ \geq 0, & \text{para } \xi_{3k|1i}^* = 0. \end{cases} \quad (3.7)$$

3.4.1. Construcción de diseños óptimos MCR para distintas funciones criterio

Calculamos ahora las iteraciones para las funciones criterio más utilizadas.

1. Diseños D -óptimos.

Las derivadas son:

$$\begin{aligned} d_{3k|1i} &= \frac{\partial \Phi}{\partial \xi_{3k|1i}} \\ &= -\tilde{\xi}_{1i} \sum_{j=1}^J \tilde{\xi}_{2j|1i,3k} \eta^t(x_1^{(i)}, x_2^{(j)}, x_3^{(k)}) M^{-1}(\xi) \eta(x_1^{(i)}, x_2^{(j)}, x_3^{(k)}). \end{aligned}$$

La función que utilizamos en el algoritmo (3.6) para este criterio es $f(d) = d$.

2. Diseños c -óptimos.

Las derivadas son:

$$d_{3k|1i} = -\tilde{\xi}_{1i} \sum_{j=1}^J \tilde{\xi}_{2j|1i,3k} [\eta^t(x_1^{(i)}, x_2^{(j)}, x_3^{(k)}) M^{-1}(\xi) c]^2.$$

De nuevo usamos la función $f(d) = d^{1/2}$ en las iteraciones para esta función criterio.

3. Diseños D_A -óptimos.

Las derivadas son:

$$\begin{aligned} d_{3k|1i} = & -\tilde{\xi}_{1i} \sum_{j=1}^J \tilde{\xi}_{2j|1i,3k} \eta^t(x_1^{(i)}, x_2^{(j)}, x_3^{(k)}) M^{-1}(\xi) A^t [AM^{-1}(\xi) A^t]^{-1} \\ & AM^{-1}(\xi) \eta(x_1^{(i)}, x_2^{(j)}, x_3^{(k)}). \end{aligned}$$

En el algoritmo (3.6) usamos la función $f(d) = d$.

El siguiente Teorema se demuestra de manera análoga al Teorema 3.1.

Teorema 3.2. *Si $E[y(\underline{x})] = \underline{\theta}_1 \underline{x}_1 + \underline{\theta}_2 \underline{x}_2 + \underline{\theta}_3 \underline{x}_3$, entonces el diseño D —óptimo marginal y condicionalmente restringido es equivalente al diseño $D_{(\theta_2, \theta_3)}$ —óptimo marginal y condicionalmente restringido.*

4. Diseños L_A —óptimos.

Las derivadas son:

$$d_{3k|1i} = -\tilde{\xi}_{1i} \sum_{j=1}^J \tilde{\xi}_{2j|1i, 3k} \eta^t(x_1^{(i)}, x_2^{(j)}, x_3^{(k)}) M^{-1}(\xi) A^t A M^{-1}(\xi) \eta(x_1^{(i)}, x_2^{(j)}, x_3^{(k)}).$$

Aplicamos el algoritmo (3.6) con $f(d) = d^{1/2}$ (Torsney [Tor83]).

En todos los ejemplos que proponemos a continuación, los diseños condicionales iniciales $\xi_{3k|1i}$ que consideramos son uniformes.

3.4.2. Ejemplos

Sea $\chi_1 = \chi_2 = \chi_3 = \{0, 1\}$ y $E[y(x)] = \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$. Supongamos que los valores de x_1 conocidos de antemano, generan el diseño

$$\tilde{\xi}_1 = \begin{Bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{Bmatrix}.$$

Además la información a priori que suponemos sobre la variable x_2 antes de que el experimento sea llevado a cabo, viene dada por:

$$\tilde{\xi}_{2|13}(\cdot | 0, 0) = \begin{Bmatrix} 0 & 1 \\ 3/4 & 1/4 \end{Bmatrix}, \quad \tilde{\xi}_{2|13}(\cdot | 1, 0) = \begin{Bmatrix} 0 & 1 \\ 2/3 & 1/3 \end{Bmatrix},$$

$$\tilde{\xi}_{2|13}(\cdot | 0, 1) = \begin{Bmatrix} 0 & 1 \\ 4/5 & 1/5 \end{Bmatrix}, \quad \tilde{\xi}_{2|13}(\cdot | 1, 1) = \begin{Bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{Bmatrix}.$$

Nuestro objetivo consiste en elegir los diseños condicionales $\xi_{3|1}(\cdot | x_1^{(i)})$ de manera óptima. Empezando el algoritmo con todos los pesos iguales, $\xi_{3k|1i} = 1/2$, y aplicando (3.6) obtenemos de manera monótona los diseños óptimos condicionales para diferentes funciones criterio como se muestra en la Tabla 4.2.

x_1	0	1	Criterio
$\xi_{3 1}^*(1 x_1)$	1	0.5435	$D-$, D_A -optimización
$\xi_{3 1}^*(1 x_1)$	1	0.5050	c -optimización
$\xi_{3 1}^*(1 x_1)$	1	0.6317	L_A -optimización

Tabla: 3.2: Diseños óptimos MCR usando un algoritmo multiplicativo

Consideramos

$$c^t = (0, 0, 1) \text{ and } A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

para $c-$, D_A- y L_A -optimización respectivamente.

De nuevo la convergencia no está asegurada por la monotonicidad, pero en la Tabla 3.3, observamos que $F_{2j|1i} \geq -0.1$ ($n = 1$) con un bajo número de iteraciones, incluso para -0.001 , ($n = 3$).

Criterio	n=1	n=2	n=3	n=4
D -optimización	3	7	16	25
D_A -optimización	3	5	8	12
c -optimización	7	11	15	19
L_A -optimización	6	20	54	104

Tabla: 3.3 : Número de iteraciones necesarias para alcanzar

$$\min_{1 \leq i \leq I, 1 \leq j \leq J} \{F_{2j|1i}\} \geq -10^{-n}$$

Con este ejemplo, confirmamos que el diseño D -óptimo MCR es equivalente a la construcción de diseños D -óptimos cuando los

parámetros de interés son el de la variable no controlable con valores conocidos antes de que el experimento sea llevado a cabo, y el parámetro de la variable controlable. Esto ilustra el Teorema 3.2.

3.5. Diseños óptimos para algunos casos reales

3.5.1. Procedimiento de densificación para la fabricación de combustible nuclear

El elemento combustible nuclear, es un conjunto de barras que contienen el uranio fisible unidas en forma de haces de manera que puedan ser manejadas de forma conjunta.

La estructura o forma de estos haces, está motivada por las funciones que deben cumplir estos elementos en el reactor. La primera es generar calor y, además, poder transmitirlo al medio líquido refrigerante, normalmente agua, por ello su geometría ha de ser tal que permite un flujo adecuado de ésta, de una manera constante y homogénea.

Para que se genere el calor, es necesario que se produzcan las reacciones nucleares adecuadas, es decir, que neutrones de una determinada energía impacten en los núcleos atómicos de uranio 235 (isótopo fisible del uranio) produciéndose la emisión de energía (calor) y más neutrones. Para que este fenómeno se produzca de una manera controlada, es decir, que no llegue a producirse la reacción en cadena, es necesario conseguir que el número de neutrones que impacta sobre los átomos de uranio sea fijo. Para conseguir esto, será necesario eliminar una gran parte de los producidos durante las reacciones, ello se consigue bien absorbiéndolos con determinadas sustancias situadas en las barras de control, o bien haciendo que se pierdan, es decir, que no lleguen a impactar con ningún otro átomo de uranio.

Para obtener lo anterior, se dispone el uranio de acuerdo en una distribución geométrica muy estudiada y calculada. De ahí se deduce el diámetro de las barras elementales, su situación espacial y las características del uranio contenido en el interior de las barras, formando pastillas de un determinado tamaño porosidad y densidad. Los poros son creados durante el proceso de calentamiento en la central nuclear. Para este propósito se añaden algunos aditivos a la mezcla inicial. Después de un tiempo en el horno, los aditivos son quemados formando los poros en las pastillas. Las variables consideradas son:

1. “Densidad inicial”, x_1 cuyos valores van desde 94.9 % a 96.7 %, y el espacio de diseño para esta variable es $\chi_1 = \{94.9, 95.0, 95.1, 95.2, 95.3, 95.4, 95.5, 95.6, 95.7, 95.8, 95.9, 96.0, 96.1, 96.2, 96.3, 96.4, 96.5, 96.6, 96.7\}$. La Tabla 3.4 muestra la distribución de frecuencias de esta variable para 392 casos.
2. “Porcentaje de aditivo U_3O_8 ”, $x_2 \in \chi_2 = \{0, 10, 20\}$.

Las propiedades de las variables son: x_1 no está sujeta al control del experimentador, pero sus valores son conocidos de antemano antes que el test sea llevado a cabo, y x_2 es una variable bajo el control del experimentador. La variable respuesta fue considerada como la densidad final (incremento de la densidad teórica). Existe interés en describir la densidad final como función de la densidad inicial y del porcentaje de aditivo. El modelo considerado en la práctica es un modelo de regresión cuadrática en las dos variables explicativas, $y(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_{12} x_1 x_2 + \theta_{11} x_1^2 + \theta_{22} x_2^2$ con las hipótesis usuales de normalidad, homocedasticidad e independencia.

Iniciando el algoritmo (3.4) con los mismos pesos iniciales para los diseños condicionales, $\xi_{2j|1i} = 1/3$, obtenemos los diseños de la Tabla 3.5. De este modo, para una densidad inicial $x_1 = 96.1$ ningún aditivo se usará en el 46 % de los casos, un 10 % en el 8 % de los casos, y un 20 % en el 46 % de los casos. Observamos que los diseños óptimos condicionales son casi simétricos y con cierto equilibrio

x_1	94.9	95.1	95.2	95.3	95.4	95.5	95.6	95.7	95.8
n_1	1	3	14	59	52	29	25	32	36

x_1	95.9	96.0	96.1	96.2	96.3	96.4	96.5	96.6	96.7
n_1	29	36	38	12	10	8	2	3	3

Tabla: 3.4: Frecuencias de porcentajes de aditivos U_3O_8 para los 392 casos considerados en el ejemplo

en los puntos medios. Por ejemplo, para $x_1 = 95.9$ los pesos son casi iguales. Pero en general, hay una preferencia por los valores extremos, como suele ocurrir en la teoría del diseño óptimo. La eficiencia del diseño usado en la práctica es del 71.5 %. Es decir, con el diseño óptimo se puede ahorrar un 30 % para obtener resultados con una precisión similar, o en otras palabras, con el mismo tamaño muestral podrían haberse obtenido mejores resultados con el diseño óptimo calculado.

De nuevo, se comprobó que el algoritmo era monótono y el número de iteraciones necesarias para conseguir que $\min_{1 \leq i \leq I, 1 \leq j \leq J} \{F_{2j|i}\} \geq -10^{-1}, -10^{-2}, -10^{-3}$ y -10^{-4} fueron 2, 4, 12 y 95 respectivamente.

3.5.2. Una prueba física para la predicción de la morbilidad quirúrgica de cáncer de pulmón

El uso de métodos científicos de investigación en medicina está siendo cada vez más frecuente. En este ejemplo, consideramos un modelo para encontrar diseños experimentales con el objeto de predecir la morbilidad quirúrgica en pacientes con diagnóstico de cáncer de pulmón sometidos a una resección pulmonar, es decir, la ocurrencia de complicaciones y/o fallecimiento durante la intervención o durante los 30 días posteriores a ésta a partir de las características

x_1	$\tilde{\xi}_1(x_1)$	$\xi_{2 1}(0 x_1)$	$\xi_{2 1}(10 x_1)$	$\xi_{2 1}(20 x_1)$
94.9	0.0026	0,4316	0,1368	0,4316
95.1	0.0077	0.4457	0.1084	0.4459
95.2	0.0357	0.4445	0.1110	0.4445
95.3	0.1505	0.4266	0.1468	0.4266
95.4	0.1327	0.3742	0.2517	0.3741
95.5	0.0740	0.2911	0.4178	0.2911
95.6	0.0638	0.2324	0.5323	0.2323
95.7	0.0816	0.2246	0.5508	0.2246
95.8	0.0918	0.2616	0.4768	0.2616
95.9	0.0740	0.3331	0.3339	0.3330
96.0	0.0918	0.4105	0.1789	0.4106
96.1	0.0969	0.4595	0.0810	0.4595
96.2	0.0306	0.4788	0.0424	0.4788
96.3	0.0255	0.4834	0.0331	0.4835
96.4	0.0204	0.4815	0.0370	0.4815
96.5	0.0051	0.4753	0.0494	0.4753
96.6	0.0077	0.4661	0.0679	0.4660
96.7	0.0077	0.4549	0.0902	0.4549

Tabla: 3.5: Diseños óptimo MCR para el procedimiento de densificación

físicas del paciente, del ejercicio físico realizado y de la existencia o no de desaturación de oxígeno en sangre durante la prueba (Varela y otros [Var01]). El ejercicio físico consiste en una prueba en bicicleta estática.

López-Fidalgo y Garcet-Rodríguez [Lóp04a], consideraron las siguientes variables principales: el “porcentaje de volumen de aire expirado”, $x_1 \in \chi_1 = \{0, 1\}$ donde cero indica una mala función respiratoria, “disminución o no de saturación de oxígeno en sangre por debajo del 90 %”, $x_2 \in \chi_2 = \{0, 1\}$ donde cero indica que la desaturación no tiene lugar y el “tiempo del ejercicio en minutos”, $x_3 \in \chi_3 = \{6, 12\}$ (tiempos estándares para la realización de la prueba). Las propiedades de estas variables son: x_1 es una variable

no controlable, pero sus valores son conocidos antes de que el ejercicio sea realizado, x_2 no está bajo el control del experimentador y sus valores son observados durante la prueba, y x_3 es una variable que está bajo el control del experimentador.

La variable respuesta fue considerada como binaria, donde 1 indica la presencia de morbilidad. El modelo considerado fue el de una regresión logística, $y \sim Bi(1, \theta(\gamma^t x))$, donde

$$\theta(\gamma^t x) = \frac{1}{1 + e^{-(\gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 x_3)}}.$$

Como el modelo es no lineal, los diseños dependen de los parámetros, de modo que necesitamos suponer valores nominales para ellos. El tiempo de ejercicio que se ha llevado a cabo, según el protocolo de Bruce (Varela y otros [Var01]), ha sido de 12 minutos. Como no hay observaciones para 6 minutos, realizamos una simulación teniendo en cuenta el experimento y la intuición de los experimentadores, así como los datos que ellos tenían hasta el momento. Un análisis de regresión logística sobre los datos simulados nos llevó a las siguientes estimaciones para los parámetros: $\gamma^t = (-1,603, 1,376, -0,276)$. Estos fueron usados como valores nominales. El conocimiento de los valores de x_1 conduce al siguiente diseño

$$\tilde{\xi}_1 = \left\{ \begin{array}{cc} 0 & 1 \\ 0.2391 & 0.7609 \end{array} \right\}.$$

Esto significa que aproximadamente el 24 % de los pacientes tienen una baja función pulmonar y el 75 % tienen una función pulmonar alta. Por otro lado, x_2 será conocida cuando la prueba sea realizada, de modo que el modelo es no lineal, los diseños dependen de los parámetros, de modo que, debemos tomar una decisión sin el conocimiento de esta variable que no se encuentra bajo nuestro control. No obstante, suponemos una información a priori sobre esta variable, x_2 , condicionada por las otras variables. La experiencia y la intuición de los profesionales proporcionó la siguiente distribución:

$$\tilde{\xi}_{2|13}(\cdot | 0, 12) = \begin{Bmatrix} 0 & 1 \\ 0.6366 & 0.3634 \end{Bmatrix}, \quad \tilde{\xi}_{2|13}(\cdot | 1, 12) = \begin{Bmatrix} 0 & 1 \\ 0.9143 & 0.0857 \end{Bmatrix},$$

que provienen de los datos que los profesionales ya tenían, y

$$\tilde{\xi}_{2|13}(\cdot | 0, 6) = \begin{Bmatrix} 0 & 1 \\ 0.7455 & 0.2545 \end{Bmatrix}, \quad \tilde{\xi}_{2|13}(\cdot | 1, 6) = \begin{Bmatrix} 0 & 1 \\ 0.6714 & 0.3286 \end{Bmatrix},$$

que se fijan a partir de la intuición y experiencia del equipo médico.

Iniciando el algoritmo (3.6) con los mismos pesos iniciales, $\xi_{3k|1i} = \frac{1}{2}$, obtenemos los diseños que aparecen en la Tabla 3.6. Consideramos $c_1^t = (1, 0, 0)$, $c_2^t = (0, 1, 0)$ y $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$. Estas elecciones se centran en el estudio de los dos primeros parámetros. No hay que olvidar que el tiempo del ejercicio se incluye en el modelo para eliminar su influencia, pues la desaturación depende del tiempo del ejercicio. El diseño usado en la práctica tiene una baja eficiencia. Por ejemplo, desde el punto de vista de los diseños D –óptimos, con el diseño óptimo que hemos calculado, solamente el 60.8 % de los pacientes serían necesarios para conseguir la misma precisión en las estimaciones. Además, los diseños que se proponen indican la realización del ejercicio para un menor tiempo en la mayoría de los casos, de modo que a los pacientes se les molesta menos con el ejercicio.

x_1	0	1	Criterio	Eficiencias (%)
$\xi_{3 1}^*(12 x_1)$	0.888	0	MCR D –optimización	60.8
$\xi_{3 1}^*(12 x_1)$	0.364	0	MCR D_A –optimización	45.1
$\xi_{3 1}^*(12 x_1)$	0.065	0	MCR L_A –optimización	47.6
$\xi_{3 1}^*(12 x_1)$	0.447	0	MCR c_1 –optimización	25.8
$\xi_{3 1}^*(12 x_1)$	0	0	MCR c_2 –optimización	71.8

Tabla: 3.6: Diseños óptimos MCR y eficiencias de los diseños usados en la práctica

La Tabla 3.7 muestra la velocidad de convergencia del algoritmo en este ejemplo. Cabe destacar que para D_A -optimización el algoritmo es más rápido que para otros criterios.

Criteria	n=1	n=2	n=3	n=4
D -optimización	4	22	63	123
D_A -optimización	3	9	25	41
L_A -optimización	18	53	102	141
c_1 -optimización	19	46	73	100
c_2 -optimización	72	200	418	708

Tabla: 3.7 : Número de iteraciones necesarias para que $\min_{1 \leq i \leq I, 1 \leq j \leq J} \{F_{2j|1i}\} \geq -10^{-n}$

3.6. Conclusiones

En este capítulo, hemos desarrollado una serie de algoritmos cuya finalidad ha sido el cálculo de diseños óptimos marginalmente restringidos y marginal y condicionalmente restringidos. Mediante estos algoritmos, hemos sido capaces de resolver algunos problemas reales como el problema de densificación y el problema de la predicción de la morbilidad quirúrgica en pacientes con cáncer de pulmón. En muchos ejemplos reales, nos encontramos con estos tipos de variables.

Las aplicaciones en la industria química, así como en otros campos están muy extendidas. Otro problema de este tipo aparece frecuentemente en el control de calidad para la materia prima proveniente de diversos proveedores y finalmente usada en un caro proceso de producción. Buenos diseños experimentales deberían ser proporcionados para predecir de manera eficiente los defectos usando una estandarización del test de la materia prima.

El tipo de materia prima, el resultado del test o la severidad del test son variables típicas que aparecen en esta clase de problemas. Aunque la primera variable no está bajo control, sus valores son conocidos antes de la realización del test. La segunda variable no está bajo el control del experimentador y sus valores son solamente observados después de que el test sea llevado a cabo. Finalmente, la tercera variable puede suponerse como una variable completamente bajo el control del experimentador.

La variable respuesta la podemos considerar como una variable binaria donde la respuesta es ser desecheda o no. Podemos considerar el modelo de regresión logística de una forma análoga al ejemplo de la prueba de esfuerzo.

Otro posible problema de este tipo podría ser el siguiente: La fragilidad del hormigón puede ser probada en el laboratorio usando tres tamaños distintos de ensayo. Además hay dos tipos de hormigón (pesado y ligero, medidos en Pascales). Después del test, el hormigón necesita pasar con éxito un proceso de control de calidad antes de ser aprobado para su uso en la construcción.

La variable respuesta es binaria, el hormigón es o no aceptado para la construcción. Tres clases de variables son consideradas como en el ejemplo de arriba. Por una parte el tipo de hormigón. Esta variable no está bajo el control del experimentador y sus valores son conocidos de antemano. El resultado del test en el laboratorio, no está bajo control y sus valores solamente son observados después de que el test sea realizado. El tamaño de la probeta de ensayo y el tiempo del test están bajo el control del experimentador. Esto introduce el caso de dos variables bajo control, reflejando un nuevo grado de dificultad que podrá ser considerado en un futuro trabajo.

Estudios computacionales muestran una rápida convergencia de los algoritmos propuestos, pero cerca del óptimo, la tasa de convergencia disminuye considerablemente (ver Tablas 3.1, 3.3, 3.7 y la sección final de 3.5.1).

Finalmente, merece la pena destacar que los Teoremas 3.1 y 3.2 muestran la equivalencia entre diseños D -óptimos MR o MCR y

los diseños D_s -óptimos. Podemos ver en el Ejemplo 1 de la sección 3.3.2 que los diseños c -óptimos y D_{θ_2} -óptimos son los mismos, ratificando de este modo el Teorema 3.1. Análogamente el Teorema 3.2 queda ilustrado por el ejemplo de la sección 3.4.2 donde el diseño D -óptimo MCR para todos los parámetros, es también el D -óptimo MCR cuando los parámetros de interés son los de la variable no controlable con valores desconocidos y el parámetro de la variable controlable.

Aunque la monotonidad no garantiza la rapidez de convergencia, una elección apropiada de la función $f(d)$ proporciona diseños casi óptimos en pocas iteraciones. Esto lo podemos observar en la Tabla 3.1 donde vemos que para diseños D -, D_A -, c - y L_A -óptimos $F_{2j|1i} \geq -0.1$ se obtiene en una o dos iteraciones.

Capítulo 4

D-Óptimos Secuenciales MR

Los **objetivos** de este capítulo son: En primer lugar resolver el problema de obtener diseños aproximados D -óptimos marginalmente restringidos cuando el diseño marginal sigue una distribución discreta y/o continua. En segundo lugar, proponer algoritmos para el cálculo de diseños cuando las observaciones van apareciendo a lo largo del estudio.

Las principales **contribuciones** son: En el apartado 4.2 se presenta mediante un ejemplo el modo de calcular el D -óptimo MR cuando el diseño marginal, ξ_1 se distribuye como una distribución discreta, en particular, como la binomial.

En el apartado 4.3 proponemos dos técnicas para calcular D -óptimos MR suponiendo, en este caso, que el diseño marginal se distribuye como una distribución continua. La primera de ellas consiste en la discretización de esta distribución, y la segunda en la modelización, mediante una familia de funciones, de los diseños que se llevan a cabo en la práctica.

En el apartado 4.4 presentamos un algoritmo secuencial, basado en el que encontramos en Fedorov [Fed72].

4.1. Introducción

En muchas situaciones experimentales algunas variables explicativas no están bajo el control del experimentador, como hemos comentado en los capítulos anteriores. Bajo este marco, diferentes aproximaciones pueden usarse cuando las observaciones llegan de manera secuencial. El objetivo principal es diseñar de manera óptima los experimentos cuando una nueva unidad experimental con un valor particular para x_1 aparece. Se consideran dos formas para diseñar:

1. Suponer una distribución particular para la variable x_1 , por ejemplo una distribución Normal o una distribución Binomial.
2. Diseñar de manera óptima para un primer conjunto de observaciones y después optimizar el diseño cada vez que una nueva unidad experimental aparezca.

Este es un problema no resuelto en la literatura. En este capítulo tratamos de dar una aproximación a la solución de este problema.

Por último, ocurre que en muchos experimentos, debido al tiempo de duración de los mismos, podemos obtener información adicional a la que teníamos cuando comenzamos. Esta información puede llegar de distintas fuentes. Lógicamente, no tiene sentido construir un diseño sin incluir esta nueva información, como tampoco tendría sentido empezar nuestra investigación cada vez que obtuviéramos una nueva información para nuestro diseño. En estos casos se usan los llamados *diseños secuenciales* que determinan nuevos diseños usando la información de las etapas previas. Para nuestro objeto de estudio, los diseños marginalmente restringidos, proponemos un algoritmo secuencial que comienza con un diseño marginal, y nuevas observaciones van apareciendo. Se supone que estas siguen una distribución de probabilidad, de modo que cuando aparece una nueva observación, el diseño marginal cambia y calculamos el mejor diseño D -óptimo MR para este diseño marginal.

4.2. D-óptimos MR con distribución marginal discreta

Consideremos el modelo de regresión a través del origen,

$$y(x) = \beta_1 x_1 + \beta_2 x_2 + \varepsilon,$$

donde $x_1 \in \chi_1 = \{0, 1, 2\}$ y $x_2 \in \chi_2 = \{0, 1\}$, y supongamos que el diseño marginal se distribuye como una binomial $\tilde{\xi}_1 \equiv \text{Bi}(2, \theta)$, es decir,

$$\tilde{\xi}_1(x_1) = \binom{2}{x_1} \theta^{x_1} (1 - \theta)^{2-x_1}, \quad x_1 = 0, 1, 2.$$

Los diseños que se llevan a cabo en la práctica, y por tanto, los que tenemos que calcular son los condicionales:

$$\xi_{2|1}(\cdot | x_1) = \begin{Bmatrix} 0 & 1 \\ 1 - p(x_1) & p(x_1) \end{Bmatrix}.$$

Sea $\theta = 1/5$. Maximizando el determinante de la matriz de información, $M(\xi) = \sum_{x_1} \sum_{x_2} f(x_1, x_2) f^t(x_1, x_2) \tilde{\xi}_1(x_1) \xi_{2|1}(\cdot | x_1)$, obtenemos:

$$p(0) = 1, \quad p(1) = \frac{3}{4} \quad \text{y} \quad p(2) = 0.$$

De modo que el diseño conjunto es:

$$\xi^* = \begin{Bmatrix} (0, 1) & (1, 0) & (1, 1) & (2, 0) \\ \frac{16}{25} & \frac{2}{25} & \frac{6}{25} & \frac{1}{25} \end{Bmatrix}.$$

Añadiendo un término constante al modelo de regresión lineal,

$$E[y(x)] = \beta_0 + \beta_1 x_1 + \beta_2 x_2,$$

los diseños óptimos condicionales son

$$p(0) = \frac{7}{16}, \quad p(1) = 1 \quad \text{and} \quad p(2) = 0,$$

obteniendo el diseño conjunto:

$$\xi^* = \begin{Bmatrix} (0, 0) & (0, 1) & (1, 1) & (2, 0) \\ \frac{9}{25} & \frac{7}{25} & \frac{8}{25} & \frac{1}{25} \end{Bmatrix}.$$

4.3. D-óptimos MR con distribución marginal continua

Bajo las mismas condiciones que en la sección anterior, consideremos que el espacio de diseño de la variable no controlable es $\chi_1 = [0, 1]$ y que el diseño marginal $\tilde{\xi}_1 \equiv \text{Beta}(\alpha, \beta)$, es decir,

$$\tilde{\xi}_1(x_1) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x_1^{\alpha-1} (1 - x_1)^{\beta-1}.$$

Los diseños que se llevan a cabo en la práctica, y por tanto en los que estamos interesados, son los diseños condicionales $\xi_{2|1}(\cdot|x_1)$.

Bajo las condiciones anteriores, la matriz de información viene dada por:

$$M(\xi) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 \begin{pmatrix} x_1^{\alpha+1}(1-x_1)^{\beta-1} & x_1^\alpha(1-x_1)^{\beta-1}p(x_1) \\ x_1^\alpha(1-x_1)^{\beta-1}p(x_1) & x_1^{\alpha-1}(1-x_1)^{\beta-1}p(x_1) \end{pmatrix} dx_1.$$

Debemos obtener la función $p(x_1)$ de modo que el determinante de dicha matriz sea máximo. Este problema no es sencillo debido a que la única restricción sobre la función $p(x_1)$ es que debe estar comprendida entre 0 y 1. Se proponen dos soluciones para este problema.

4.3.1. Discretización de la distribución de probabilidad del diseño marginal

La distribución de probabilidad continua será aproximada por una distribución discreta con soporte finito. Dividiendo el espacio de diseño $\chi_1 = [a, b]$ en l partes de la misma probabilidad, consideramos una distribución marginal discreta. Sean $\{z_1, \dots, z_{l-1}\}$ los correspondientes cuantiles, es decir, $F(z_i) = \frac{i}{l}$. Los puntos del soporte del diseño marginal, serán los puntos medios de cada intervalo de

acuerdo a la distribución continua de origen. Se incluirán además los puntos extremos, pues como es conocido, el diseño óptimo con frecuencia recurre a ellos. Además serán necesarios para la interpolación que se hará posteriormente con el diseño óptimo obtenido:

$$\begin{aligned} t_0 &= a, \\ t_i &= \int_{z_{i-1}}^{z_i} \frac{t\tilde{\xi}_1(t)}{F(z_i)-F(z_{i-1})} dt = l \int_{z_{i-1}}^{z_i} t\tilde{\xi}_1(t), \quad i = 1, \dots, l, \\ t_{l+1} &= b, \end{aligned}$$

donde $\tilde{\xi}_1(t)$ es la función de densidad, F es la función de distribución de $\tilde{\xi}_1$, $z_0 = a$ y $z_l = b$. Entonces tendremos el diseño marginal discreto siguiente:

$$\tilde{\xi}_1^{(l)} = \left\{ \begin{array}{cccc} t_0 & t_1 & \dots & t_{l+1} \\ \frac{1}{l+2} & \frac{1}{l+2} & \dots & \frac{1}{l+2} \end{array} \right\},$$

y con este, calculamos los diseños óptimos condicionales como se hizo en la sección anterior:

$$\xi_{2|1}^{(l)}(\cdot|t_i), \quad i = 0, \dots, l+1.$$

Si se obtiene una nueva observación, $x_1 \in \chi_1$, se puede calcular $\xi_{2|1}(\cdot|x_1)$ utilizando interpolación lineal como sigue:

$$\xi_{2|1}(\cdot|x_1) = \xi_{2|1}^{(l)}(\cdot|t_i) + \frac{\xi_{2|1}^{(l)}(\cdot|t_{i+1}) - \xi_{2|1}^{(l)}(\cdot|t_i)}{t_{i+1} - t_i}(x_1 - t_i),$$

$$i = 0, \dots, l+1, \quad x_1 \in [t_i, t_{i+1}]$$

es decir,

$$p(x) = p(t_i) + \frac{p(t_{i+1}) - p(t_i)}{t_{i+1} - t_i}(x - t_i), \quad i = 0, \dots, l.$$

Ahora, necesitamos una cota inferior de la eficiencia. Usamos la cota proporcionada por López-Fidalgo y Garcet-Rodríguez (2004):

$$1 + \frac{\int_{\chi_1} \inf_{x_2 \in \chi_2} \partial \Phi[M(\xi^{(D)}), M_{(x_1, x_2)}] \tilde{\xi}_1(dx_1)}{\Phi[M(\xi^{(D)})]}.$$

En el caso de D -optimización, en el capítulo 2 vimos que la cota anterior la podíamos escribir del siguiente modo:

$$\text{Eff}_{MRD}(\xi) \geq 2 - \frac{1}{2} \int_0^1 \max_{x_2 \in \{0,1\}} f^t(x) M^{-1}(\xi^{(D)}) f(x) \tilde{\xi}_1(dx_1). \quad (4.1)$$

Ejemplo

Supóngase que $\tilde{\xi}_1$ sigue una distribución beta, con parámetros $(1, 2)$. Si el intervalo $[0, 1]$ se divide en l partes, entonces:

$$\begin{aligned} z_i &= 1 - \sqrt{\frac{l-i}{l}}, \quad i = 0, 1, \dots, l, \\ t_0 &= 0, \\ t_i &= 2l \left[\frac{z_i^2 - z_{i-1}^2}{2} + \frac{z_{i-1}^3 - z_i^3}{3} \right] \quad i = 1, \dots, l, \\ t_{l+1} &= 1. \end{aligned}$$

Notemos $p_i = \xi_{2|1}^{(l)}(1 \mid t_i)$, $i = 0, \dots, l+1$. Por tanto, el determinante de la matriz de información para el diseño discretizado es

$$I_D(\xi^{(l)}) = \det M(\xi^{(l)}) = \frac{1}{(l+2)^2} \left[\left(1 + \sum_{r=1}^l t_r^2 \right) \left(\sum_{r=0}^{l+1} p_r \right) - \left(\sum_{r=1}^{l+1} p_r t_r \right)^2 \right],$$

que es una función creciente en p_0 , de modo que el valor óptimo lo alcanzará para $p_0^* = 1$. Por otro lado, las derivadas respecto a p_i , $i = 1, \dots, l+1$ son

$$\frac{\partial I_D(\xi^{(l)})}{\partial p_i} = \frac{1}{(l+2)^2} \left[\left(1 + \sum_{r=1}^l t_r^2 \right) - 2t_i \left(\sum_{r=1}^{l+1} p_r t_r \right) \right] \quad i = 1, \dots, l+1.$$

Sólo una de estas derivadas puede ser cero. El resto, tienen que ser siempre positivas o negativas, de modo que $p_i^* = 0$ o 1, quizás con la excepción de algunos valores $p_{i_0}^*$. Este valor viene dado por

$$p_{i_0}^* = \frac{\sum_{r=0, r \neq i_0}^{l+1} t_r^2 - t_{i_0} \sum_{r=1}^{i_0-1} t_r}{t_{i_0}}. \quad (4.2)$$

Como $t_i < t_{i+1}$, si la derivada es negativa para i también lo es para la sucesiva $i + 1$. Por tanto, hay tres casos posibles para los pesos del diseño óptimo,

1. Si todas las derivadas son positivas entonces el diseño óptimo tiene los siguientes pesos $\{1, 1, \dots, 1\}$.
2. Si una de las derivadas se anula, entonces el diseño óptimo tiene pesos $\{1, \dots, 1, p_{i_0}, 0, \dots, 0\}$.
3. Si todas las derivadas son negativas, entonces el diseño óptimo tiene pesos $\{1, 0, \dots, 0\}$.

Para determinar cuál de los tres casos anteriores es el óptimo, podemos usar el siguiente procedimiento,

1. Calcular $I_D(\xi^{(l)})$ para todas las posibilidades $\{1, \dots, 1, 0, \dots, 0\}$.
2. Elegir el máximo de ellos. Sea i_0 el subíndice para el último 1.
3. Calcular p_{i_0} y p_{i_0+1} con la fórmula (4.2).
4. Calcular $I_D(\xi^{(l)})$ para $\{1, \dots, 1, p_{i_0}, 0, \dots, 0\}$ y $\{1, \dots, 1, p_{i_0+1}, 0, \dots, 0\}$.

Después de la interpolación y usando la distribución Beta, la matriz de información para el diseño final es:

$$M(\xi) = 2 \begin{pmatrix} 1/12 & \sum_{i=0}^l \int_{t_i}^{t_{i+1}} p^{(D)}(x_1) x_1 (1 - x_1) dx_1 \\ \sum_{i=0}^l \int_{t_i}^{t_{i+1}} p^{(D)}(x_1) x_1 (1 - x_1) dx_1 & \sum_{i=0}^l \int_{t_i}^{t_{i+1}} p^{(D)}(x_1) (1 - x_1) dx_1 \end{pmatrix}.$$

Dividamos el intervalo $[0, 1]$ en dos partes equiprobables, $z_0 = 0$, $z_1 = 0.29$ y $z_2 = 1$, de modo que los puntos del soporte del diseño marginal son: $t_0 = 0$, $t_1 = 0.13$, $t_2 = 0.52$ y $t_3 = 1$, y consideremos un diseño marginal con los mismos pesos,

$$\tilde{\xi}_1 = \begin{Bmatrix} 0 & 0.13 & 0.52 & 1 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{Bmatrix}.$$

Bajo las condiciones anteriores, los diseños condicionales, $p_i = \xi_{2|1}(1|t_i)$, pueden ser calculados como en la sección anterior (véase Tabla 4.1). Aplicando (4.1) a este caso, tenemos la cota inferior para la eficiencia mostrada en la Tabla 4.1.

Repetimos el proceso para $l = 3, 4, 8$ y 10 . En la Tabla 4.1 se presentan los diseños óptimos condicionales y las eficiencias obtenidos para cada discretización.

l	i	1	2	3	4	5	6	7	8	9	10	Cota	Eff
2	t_i	0.13	0.52										
	p_i	1	1									72.68	89.72
3	t_i	0.086	0.29	0.62									
	p_i	1	1	1								60.32	86.10
4	t_i	0.065	0.21	0.39	0.67								
	p_i	1	1	1	1							53.49	84.25
8	t_i	0.032	0.099	0.17	0.25	0.34	0.44	0.57	0.76				
	p_i	1	1	1	1	1	1	1	0			77.00	91.43
10	t_i	0.025	0.078	0.13	0.19	0.25	0.33	0.41	0.50	0.61	0.79		
	p_i	1	1	1	1	1	1	1	1	0.396	0	86.36	93.33

Tabla: 4.1: Diseños óptimos condicionales, cotas para las eficiencias y eficiencias exactas en porcentajes para el ejemplo ($t_0 = 0$, $p_0 = 1$, $t_{l+1} = 1$, $p_{l+1} = 0$ en todos los casos)

Después de los pasos anteriores, podemos ver que un diseño de la forma

$$p(x_1) = \begin{cases} 1, & \text{si } x_1 \leq z \\ 0, & \text{si } x_1 > z \end{cases}$$

puede llevar al diseño óptimo. De modo que,

$$\det\{M(\xi)\} = 4 \left(\frac{z}{12} - \frac{z^2}{24} - \frac{z^4}{4} - \frac{z^6}{9} + \frac{z^5}{3} \right)$$

y el máximo se alcanza en $z = 0.5$.

La eficiencia de este diseño resulta ser 1, por tanto, hemos encontrado un diseño óptimo. De este modo, podemos calcular las eficiencias exactas para los diseños calculados anteriormente (ver Tabla 4.1).

4.3.2. Modelización de $p(x_1)$ por una familia paramétrica de funciones

Supóngase que $p(x_1)$ pertenece a una familia de polinomios,

$$p(x_1) = b_0 + b_1x_1 + \dots + b_kx_1^k,$$

donde b_0, \dots, b_k han de asegurar que $p(x_1) \in [0, 1]$ para todos los puntos de diseño, $x_1 \in \chi_1$. Se calcula la matriz de información y se maximiza el determinante del modo habitual.

Consideremos de nuevo el modelo de regresión lineal

$$E[y(x)] = \beta_1x_1 + \beta_2x_2,$$

y supongamos en particular que $p(x_1) = \gamma + (\delta - \gamma)x_1$ con $0 \leq \gamma, \delta \leq 1$. Maximizando el determinante de la matriz de información se obtiene $\gamma = 1$ y $\delta = 0$, de modo que

$$p(x_1) = 1 - x_1.$$

La cota inferior para la eficiencia es entonces:

$$2 - \frac{1}{m} \int_{\chi_1} \max_{x_2 \in \chi_2} f^t(x) M^{-1}(\xi) f(x) \tilde{\xi}_1(dx_1) =$$

$$= 2 - \frac{1}{2} \int_0^1 \max_{x_2 \in \chi_2} (x_1 \ x_2) \begin{pmatrix} 8 & -2 \\ -2 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \frac{\Gamma(3)}{\Gamma(1)\Gamma(2)} (1-x_1) dx_1 = 0.916.$$

Modelizando $p(x_1)$ por un polinomio de orden 2, $p(x_1) = b_0 + b_1x_1 + b_2x_1^2$, y maximizando el determinante de $M(\xi)$, obtenemos que $p(x_1) = 1 - \frac{3}{8}x_1 - \frac{5}{8}x_1^2$, y la cota inferior para la eficiencia es 0.905.

Las cotas obtenidas con este diseño mediante este procedimiento son bastante buenas, sin embargo, en otros casos podría ser más difícil obtener cotas tan altas. Así, considerando el modelo

$$E[y(x)] = \beta_0 + \beta_1x_1 + \beta_2x_2,$$

la matriz de información en este caso será:

$$M(\xi) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 \begin{pmatrix} x_1^{\alpha-1}(1-x_1)^{\beta-1} & x_1^{\alpha}(1-x_1)^{\beta-1} & x_1^{\alpha-1}(1-x_1)^{\beta-1}p(x_1) \\ x_1^{\alpha}(1-x_1)^{\beta-1} & x_1^{\alpha+1}(1-x_1)^{\beta-1} & x_1^{\alpha}(1-x_1)^{\beta-1}p(x_1) \\ x_1^{\alpha-1}(1-x_1)^{\beta-1}p(x_1) & x_1^{\alpha}(1-x_1)^{\beta-1}p(x_1) & x_1^{\alpha-1}(1-x_1)^{\beta-1}p^2(x_1) \end{pmatrix} dx_1.$$

y suponiendo que $p(x_1) = \gamma + (\delta - \gamma)x_1$, obtenemos una cota inferior para la eficiencia de 0.5. La misma cota se obtuvo cuando se consideró $p(x_1) = b_0 + b_1x_1 + b_2x_1^2$.

4.4. Construcción Secuencial de diseños D-óptimos MR

Se establece un diseño inicial, $\tilde{\xi}_1^{(0)}$. A partir de este, se procede a calcular el D -óptimo MR, $\xi_{MR}^{(0)}$.

Supongamos que se obtiene una nueva observación $x_1^{(1)}$, por ejemplo un nuevo paciente que se suma al estudio. Por tanto, $\tilde{\xi}_1^{(0)}$ se

convertiría en un nuevo diseño marginal $\tilde{\xi}_1^{(1)}$. Para este “nuevo” diseño marginal, podemos calcular el diseño D –óptimo restringido $\xi_{MR}^{(1)}$. Este diseño no tiene por qué basarse en el diseño anterior $\xi_{MR}^{(0)}$. Calcularemos un diseño óptimo que conserve el diseño $\xi_{MR}^{(0)}$, y elegimos el punto $x_2^{(1)}$, maximizando el determinante de la matriz de información, del siguiente modo (véase Fedorov [Fed72]):

$$x_2^{(1)} = \arg \max_{x_2} d(x_1^{(1)}, x_2, \xi_{MR}^{(0)}).$$

Lo anterior equivale a maximizar el determinante de la matriz de información.

Añadiendo el punto $(x_1^{(1)}, x_2^{(1)})$ se obtiene un nuevo diseño $\xi^{(1)}$. Continuando con este proceso, obtenemos un procedimiento secuencial para diseñar el experimento en pasos sucesivos. Al mismo tiempo, de cada diseño marginal $\tilde{\xi}_1^{(N)}$, podemos calcular el diseño MR D –óptimo $\xi_{MR}^{(N)}$. Este diseño no ha sido aplicado, pues el diseño que se ha usado ha sido $\xi^{(N-1)}$, y estos no deben ser necesariamente consecuentes uno con otro. Esta situación se ilustra en la Figura 4.1.

Lo ideal sería que ambos diseños fueran iguales, al menos de manera asintótica. No obstante, no hay garantía de este resultado en general. Hablar de convergencia de este procedimiento no tiene sentido, estrictamente hablando, puesto que el diseño obtenido en cada paso N , es el mejor para el diseño conjunto que se tiene en ese momento como resultado de incluir el nuevo punto $x_1^{(N)}$ que acaba de aparecer en el estudio. Sin embargo, se puede plantear la cuestión de cuáles son las condiciones bajo las que $\xi_{MR} = \xi$, (Figura 4.1), suponiendo que la secuencia de diseños marginales $\tilde{\xi}_1^{(N)}$ converge a $\tilde{\xi}_1$.

4.4.1. Un ejemplo real

Proponemos una modificación del ejemplo propuesto por Varela y otros [Var01], en el que un test de ejercicio podría ayudar en la evaluación del riesgo preoperativo en la cirugía de pulmón. La

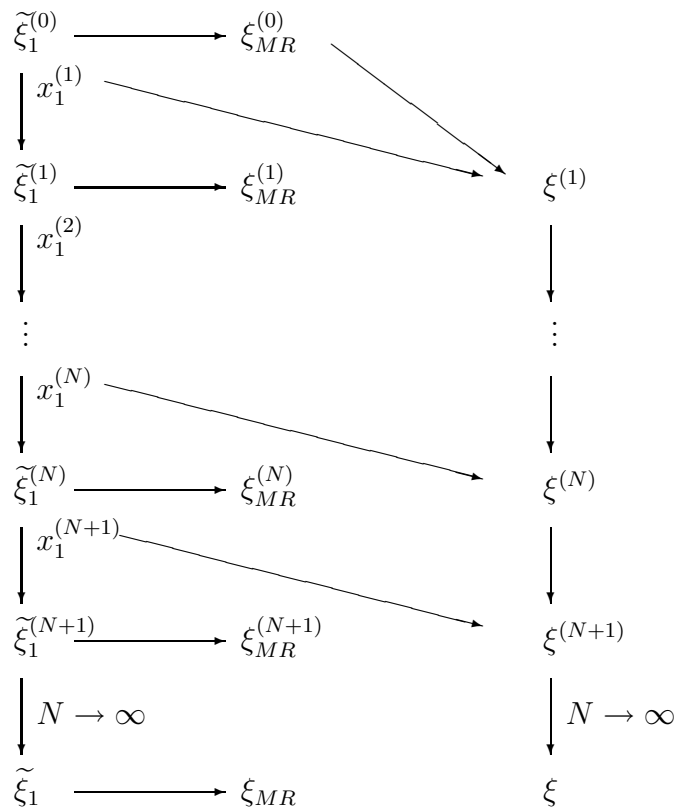


Figura 4.1 : Construcción de diseños secuenciales

principales variables consideradas en este estudio son el “porcentaje de máximo volumen expirado por el paciente” (x_1) y el “tiempo de ejercicio en minutos” (x_2). La variable x_2 está completamente bajo el control del experimentador, en el sentido de que el tiempo de ejercicio para cada paciente podría ser determinado mediante este procedimiento. La variable x_1 no está sujeta bajo control, pero sus valores son conocidos antes de realizar la prueba de ejercicio.

Llevaremos a cabo el experimento y obtenemos los resultados para un grupo de pacientes. Supongamos ahora que nuevos pacientes acuden de manera secuencial. Para cada nuevo paciente, el “porcentaje de máximo volumen de aire expirado” podría ser medido.

Ahora se necesitará tomar una decisión sobre el “tiempo de ejercicio” que debería ser seleccionado para este “nuevo” paciente específico con el objeto de optimizar el diseño.

Un ejemplo simple ilustrará este procedimiento. Una vez mas, consideramos el modelo lineal:

$$E[y(x)] = \gamma_1 x_1 + \gamma_2 x_2, \quad \chi_1 = \{0, 1\} \quad \chi_2 = \{0, 1\}.$$

Supongamos que en el paso N el diseño marginal es:

$$\tilde{\xi}_1^{(N)} = \left\{ \begin{array}{cc} 0 & 1 \\ \frac{\mu}{n} & \frac{n-\mu}{n} \end{array} \right\},$$

de modo que los diseños condicionales serán:

$$\xi_{2|1}^{(N)}(\cdot|0) = \left\{ \begin{array}{cc} 0 & 1 \\ \frac{\mu-s}{\mu} & \frac{s}{\mu} \end{array} \right\}, \quad \xi_{2|1}^{(N)}(\cdot|1) = \left\{ \begin{array}{cc} 0 & 1 \\ \frac{n-\mu-t}{n-\mu} & \frac{t}{n-\mu} \end{array} \right\},$$

con $n, \mu, s, t \in \mathbf{Z}$.

Se construyen diseños óptimos secuenciales de acuerdo a los diseños marginales que se van obteniendo. Dependiendo del valor de x_1 , tendremos dos posibilidades para cada caso.

Se puede resumir el algoritmo de obtención de diseños óptimos MR secuenciales como sigue:

1. Sea $\tilde{\xi}_1^{(0)}$ el diseño marginal inicial.
2. Calculamos $\xi_{MR}^{(0)}$ bajo la restricción del diseño marginal $\tilde{\xi}_1^{(0)}$.
3. En el paso $N + 1$ se dan dos posibilidades para el nuevo paciente que acude al hospital,

a) Si $x_1^{(N+1)} = 1$, entonces

$$x_2^{(N+1)} = \left\{ \begin{array}{ll} 1 & \text{si } \xi_{2|1}^{(N)}(1 | 1) < \frac{1}{2} \\ 0 & \text{si } \xi_{2|1}^{(N)}(1 | 1) > \frac{1}{2} \\ \text{se aleatorizará } \{0, 1\} & \text{si } \xi_{2|1}^{(N)}(1 | 1) = \frac{1}{2} \end{array} \right.$$

b) Si $x_1^{(N+1)} = 0$, entonces

$$x_2^{(N+1)} = \begin{cases} 1 & \text{si } \tilde{\xi}_1^{(N)}(1) > 0 \\ \text{se aleatorizará } \{0, 1\} & \text{si } \tilde{\xi}_1^{(N)}(1) = 0 \end{cases}$$

Apliquemos el algoritmo al ejemplo para clarificar estas ideas. Sea el diseño marginal inicial siguiente:

$$\tilde{\xi}_1^{(0)} = \begin{Bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{Bmatrix}.$$

Después de algunos cálculos se obtiene el diseño óptimo MR:

$$\xi_{MR}^{(0)} = \begin{Bmatrix} (0, 0) & (0, 1) & (1, 0) & (1, 1) \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{Bmatrix}.$$

Supongamos que por estudios retrospectivos se sabe que a largo plazo el diseño marginal tiende a seguir una distribución binomial de parámetros $n = 1$ y $\theta = 1/5$. Simulamos 8 observaciones para x_1 , por ejemplo, lanzando una moneda:

$$x_1 = 0, 0, 0, 1, 0, 0, 0, 0.$$

Aplicando el algoritmo, obtenemos los resultados dados en la Tabla 4.2. Se alcanza una eficiencia del 100 % a partir del cuarto paciente, observando una convergencia rápida en este caso.

N	$\xi_1^{(N)}$	$\xi_{MR}^{(N)}$	$\xi^{(1)}$	Eff
1	2/3, 1/3	0, 2/3, 1/6, 1/6	0, 2/3, 0, 1/3	88,8 %
2	3/4, 1/4	0, 3/4, 1/8, 1/8	0, 3/4, 0, 1/4	93,6 %
3	4/5, 1/5	0, 4/5, 1/10, 1/10	0, 4/5, 0, 1/5	94,1 %
4	4/6, 2/6	0, 4/6, 1/6, 1/6	0, 4/6, 1/6, 1/6	100 %
5	5/7, 2/7	0, 5/7, 1/7, 1/7	0, 5/7, 1/7, 1/7	100 %
6	6/8, 2/8	0, 6/8, 1/8, 1/8	0, 6/8, 1/8, 1/8	100 %
7	7/9, 2/9	0, 7/9, 1/9, 1/9	0, 7/9, 1/9, 1/9	100 %
8	8/10, 2/10	0, 8/10, 1/10, 1/10	0, 8/10, 1/10, 1/10	100 %

Tabla: 4.2: Pesos del diseño secuencial en los puntos $\{0, 1\}$ en la primera columna, y en los puntos $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$ para las segunda y tercera columnas. La última columna muestra las eficiencias del diseño secuencial respecto del correspondiente D -óptimo MR.

4.5. Conclusiones

En este capítulo se han proporcionado diseños D -óptimos cuando uno de los factores no puede estar bajo el control del experimentador, pero sus valores son conocidos antes de que el experimento sea llevado a cabo, y otro factor está bajo el control del experimentador. Los resultados de las secciones 4.2 y 4.3 pueden ser fácilmente generalizados cuando una nueva variable no controlable se observa durante el ejercicio.

El algoritmo secuencial propuesto fue motivado por el ejemplo que aparece en López-Fidalgo y Garcet-Rodríguez, [Lóp04a]. Ellos proporcionaron un diseño óptimo para predecir la morbilidad quirúrgica después de una resección pulmonar. Las principales variables consideradas fueron el “porcentaje de máximo volumen expirado por el paciente”, “la desaturación de oxígeno durante la prueba” y el “tiempo de ejercicio en minutos”. Nosotros consideramos un problema más simplificado teniendo en cuenta solamente la primera y tercera variables. Suponiendo que hemos calculado nuestro diseño, pero que un nuevo paciente llega, el diseño marginal cambia. Nuestro objetivo consiste en asignar un tiempo de ejercicio de for-

ma que el diseño conjunto sea óptimo. Los diseños secuenciales se construyen de tal modo que toda la información que se tiene en las etapas previas, es usada para sugerir cómo calcular la siguiente iteración.

Capítulo 5

D —óptimos MR con observaciones correladas

*El **objetivo** de este capítulo consiste en estudiar los diseños para modelos en los que aparecen variables explicativas no controlables, y a su vez que las respuestas puedan ser correladas. Además trataremos de comprobar que no sólo las distintas estructuras de covarianza afectan al cálculo de los diseños, sino también los parámetros que sean considerados de interés.*

*Las principales **contribuciones** de este capítulo son: En el apartado 5.3 se hace un estudio original sobre las funciones de correlación triangular, obteniendo las restricciones sobre el parámetro de modo que la matriz de covarianzas esté bien definida. En el apartado 5.4 proporcionamos el cálculo de distintos diseños marginalmente restringidos con distintas estructuras de covarianza. En los apartados 5.5, 5.6 y 5.7 planteamos y resolvemos problemas reales, dentro de este marco de diseños restringidos con observaciones correladas, adaptando un algoritmo iterativo para resolver dichos problemas. Finalmente, en el apartado 5.8 presentamos un análisis de sensibilidad, analizando las “consecuencias” de considerar distintas funciones de covarianza de las verdaderas y/o va-*

lores de los parámetros erróneos.

5.1. Introducción

La gran mayoría de los trabajos de diseño óptimo suponen que los errores son independientes y que es posible realizar réplicas. Estos resultados muestran que en vez de calcular un diseño exacto, podemos usar la llamada medida de diseño, que es proporcional al número de réplicas independientes. La información obtenida del experimento se expresa mediante la matriz de información, que como hemos visto en el capítulo 1, es una función de la medida de diseño. Sin embargo, hay muchos problemas donde la modelización con observaciones correladas es más adecuada, por ejemplo en minería, oceanografía, meteorología, reacciones químicas... Desde la teoría de diseños óptimos esto es un gran desafío que no ha sido muy tratado. Podemos encontrar referencias en Brinkulov y otros [Bri80], Abt y Welch [Abt98], Cressie [Cre93], Pázman [Paz04] o Uncinski y Atkinson [Uci04].

Bajo la hipótesis de que las observaciones puedan ser correladas, el marco teórico es muy distinto al expuesto en los primeros capítulos. Así, consideremos el modelo

$$y(x) = \sum_{i=1}^m f_i(x)\theta_i + \varepsilon(x) = f^t(x)\theta + \varepsilon(x),$$

donde $\varepsilon \sim N_n(0, \Sigma)$, siendo la matriz de covarianzas $\Sigma = (C(x, z))_{x,z}$ simétrica y definida positiva. El estimador mínimo cuadrático viene dado por

$$\hat{\theta} = (X^t \Sigma^{-1} X)^{-1} X^t \Sigma^{-1} Y,$$

y su matriz de covarianzas

$$Cov(\hat{\theta}) = (X^t \Sigma^{-1} X)^{-1}.$$

La matrix de información correspondiente a un diseño exacto V_N de N puntos es:

$$M(V_N) = \sum_{x \in \chi} \sum_{z \in \chi} f(x) \Sigma_{x,z}^{-1} f(z).$$

Bajo estas condiciones no se permite la generalización de las medidas de diseño según la teoría de Kiefer. Las matrices de información no satisfacen las propiedades de convexidad, que se verifican en las matrices de información cuando las observaciones son incorreladas. Así no tendremos el teorema de equivalencia ni otras propiedades basadas en la convexidad de los conjuntos de las matrices de información. (Para una mayor discusión véase Nachtsheim [Nac89] páginas 74 – 75).

Análogamente, los estudios realizados sobre diseño óptimo con variables no controlables restringen su atención a observaciones incorreladas, siendo esta memoria, el primer trabajo (que conozca el autor) donde se tratan los dos aspectos al mismo tiempo, es decir, observaciones correladas con variables no controlables.

5.2. Planteamiento del problema

Consideremos el modelo $E[y(x)] = \eta(s, t, \underline{\theta}_1)$, donde (s, t) pertenecen a algún compacto χ , siendo $\chi = \chi_1 \times \chi_2$ el espacio de diseño y sea Ξ el conjunto de todos los diseños en este espacio producto. Como vimos en el capítulo 2, un diseño conjunto ξ , queda completamente determinado por un diseño marginal ξ_1 sobre χ_1 y los diseños condicionales $\xi_{2|1}(\cdot|s)$ sobre χ_2 dada $s \in \chi_1$.

Si los valores de la variable s son conocidos de antemano, estos determinan un diseño marginal específico $\tilde{\xi}_1$. De este modo, el espacio Ξ_R se restringe a aquellos diseños conjuntos que tienen a $\tilde{\xi}_1$ como diseño marginal.

Supongamos que las observaciones son correladas, de modo que $\text{cov}(y_i, y_j) = C(|s_i - s_j|, |t_i - t_j|, \theta_2)$ y que pueden ser obtenidas con

errores gaussianos con media cero y varianzas finitas, de este modo observamos un proceso aleatorio estacionario gaussiano (ver Adler y Taylor [Adl05] para las condiciones de regularidad exigidas). En un principio, podemos considerar dos conjuntos de parámetros de interés: Los que describen la tendencia, $\underline{\theta}_1$ y los que caracterizan a la función de covarianza, $\underline{\theta}_2$. Para este último conjunto tenemos otras dos enfoques para calcular el diseño, que son: considerar el conjunto de parámetros, $\underline{\theta}_2$, como parámetros con un valor fijo, conocidos o considerarlos como parámetros de interés. Mediante algunos ejemplos, mostraremos la diferencia entre las consideraciones anteriores. Para la primera aproximación, la inversa de la matriz de información de Fisher tiene el significado estándar, es decir, asintóticamente es proporcional a la matriz de covarianzas de los estimadores. Sin embargo, mediante el segundo enfoque la optimización basada en la matriz de información de Fisher puede no resultar conveniente. En cualquier caso, como decíamos en la introducción, la teoría de los diseños aproximados no puede usarse en este contexto, es por esto que debemos considerar diseños exactos. Debido a que bajo este marco no existe un teorema de equivalencia para poder demostrar la convergencia de algoritmos, han sido desarrolladas otras aproximaciones. Por ejemplo, el algoritmo dado por Brimkulov y otros [Bri80]. Aunque no hay una prueba de la convergencia para este algoritmo, su eficacia ha sido mostrada empíricamente (véase Uciński y Atkinson [Uci04]).

Supongamos que estamos interesados en estimar los parámetros de la media, $\underline{\theta}_1 = (\theta_1^{(1)}, \dots, \theta_1^{(m)})$. La matriz de información de Fisher sería:

$$M_{\underline{\theta}_1} = -E \begin{pmatrix} \frac{\partial^2 L}{\partial \theta_1^{(1)2}} & \cdots & \frac{\partial^2 L}{\partial \theta_1^{(1)} \partial \theta_1^{(n)}} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 L}{\partial \theta_1^{(n)} \partial \theta_1^{(1)}} & \cdots & \frac{\partial^2 L}{\partial \theta_1^{(n)2}} \end{pmatrix},$$

siendo L el log de la función de verosimilitud. Bajo condiciones de normalidad $L = \text{const.} - \frac{1}{2} \ln \det \sigma^2 \Sigma(\theta_2) - \frac{1}{2\sigma^2} e^t \Sigma^{-1}(\theta_2) e$, donde $\Sigma(\theta_2)$ es la matriz de covarianzas correspondiente a la función de covarianza dada y e los residuos.

Suponiendo $\sigma^2 = 1$ y $X = \{f_j(s_i, t_i)\} = \{\partial\eta(s_i, t_i, \underline{\theta}_1)/\partial\theta_1^{(j)}\}$, la matriz de información será: $M = X^t \Sigma^{-1} X$, de modo que usando el criterio de D -optimización, la función a maximizar es:

$$Inf_1 = |X^t \Sigma^{-1} X|.$$

Si estamos interesados en estimar todos los parámetros, suponiendo de nuevo que $\sigma^2 = 1$ y $X = \{f_j(s_i, t_i)\} = \{\partial\eta(s_i, t_i, \theta_1)/\partial\theta_1^{(j)}\}$, la matriz de información será:

$$M = \frac{1}{2} \begin{pmatrix} 2X^T \Sigma^{-1} X & 0 \\ 0 & \text{tr}(\Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_2} \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_2}) \end{pmatrix},$$

y la función a maximizar:

$$Inf_2 = |M| = |X^t \Sigma^{-1} X| \text{tr} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_2} \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_2} \right).$$

En la Sección 5.3 serán calculados diseños D -óptimos MR para un modelo de regresión lineal y una estructura de covarianza exponencial/triangular. En las Secciones 5.4 y 5.5 resolveremos dos casos reales.

5.3. Estructura de la covarianza

El efecto de emplear una función de covarianza parametrizada u otra, juega un papel importante en el comportamiento del diseño. En este capítulo consideramos como funciones de covarianza subclases de funciones Gaussianas y Exponenciales de la clase de Matérn (Matérn [Mat86]) y clases triangulares. Además, introducimos la función de covarianza potencial modificada que permite estimar la influencia de la variable no controlable en la dependencia espacio-tiempo, y una función de covarianza algo particular que encontramos en Stein [Ste05b].

La clase de Matérn viene motivada por el amplio rango de comportamientos que cubre y la interpretabilidad de los parámetros. Existen distintas parametrizaciones de la familia de Matérn. En esta memoria elegimos la dada por Handcock y Wallis [Han94],

$$\text{cov}(u, \theta_2, v, \tau^2) = \tau^2 \delta(u) + \frac{1}{2^{v-1} \Gamma(v)} \left(\frac{2u\sqrt{v}}{\theta_2} \right)^v \mathcal{K}_v \left(\frac{2u\sqrt{v}}{\theta_2} \right),$$

donde θ_2 y v son los parámetros, u es una distancia, generalmente la distancia entre dos puntos experimentales, τ^2 es el llamado efecto nugget, es decir, el efecto debido a la presencia de microvarianza (véase Cressie [Cre93]), $\delta(u)$ es la función de Kronecker y \mathcal{K}_v es la función de Bessel modificada de tercera clase y orden v (ver Abramowitz y Stegun [Abr65], sec. 9). El parámetro v controla la diferenciabilidad del proceso y θ_2 mide la rapidez del descenso de las correlaciones según la distancia (*rango del parámetro* en la literatura de geoestadística). Esta familia incluye la *clase exponencial* $\exp(-\theta_2 u)$ (como un caso especial cuando $v=0.5$) y la *familia Gaussiana* $\exp(-\theta_2 u^2)$ de funciones de correlación (como un caso límite cuando $v \rightarrow \infty$).

Como se dijo, consideramos también las *funciones potenciales* descritas en Currin y otros [Cur91], $\exp(-\theta_2 |t - t'|^p)$ para $\theta_2 > 0$ y $0 < p \leq 2$. Gracias a la ya definida correlación producto, podemos definir la función de correlación sobre $[0, 1]^2$:

$$\exp(-\theta_{21} |s - s'|^p - \theta_{22} |t - t'|^p). \quad (5.1)$$

También, trabajaremos con las *funciones de correlación triangular*, que introducimos a continuación.

Lema 5.1. *Consideremos un modelo general con variables explicativas s y t en los intervalos χ_1 y χ_2 respectivamente, y las funciones de covarianza:*

1. $\text{cov}(y, y') = \sigma^2(1 - \theta_2(|s - s'| + |t - t'|)),$
2. $\text{cov}(y, y') = \sigma^2(1 - \theta_2|s - s'|)(1 - \theta_2|t - t'|).$

Esta estructura es coherente, es decir, la matriz de covarianzas es definida no negativa, si y solo si

1. $0 < \theta_2 < 2/(\lambda(\chi_1) + \lambda(\chi_2))$,
2. $0 < \theta_2 < 2/\max\{\lambda(\chi_1), \lambda(\chi_2)\}$

siendo λ la longitud del intervalo.

Obtendremos una correlación positiva intercambiando el 2 por un 1 en ambas cotas.

Demostración.

Sin pérdida de generalidad, supongamos que $t_1 < \dots < t_n$. Sea P_n el conjunto de las permutaciones de $\{1, \dots, n\}$, y Σ_n la matriz de covarianzas. Supongamos que $s_{\psi(1)} < \dots < s_{\psi(n)}$ para alguna permutación $\psi \in P_n$. Usando Fuzhen Zhang [Zha99] obtenemos la expresión $|\Sigma_n| = f_n(\theta_2) \prod_{1 \leq i < j \leq n} (x_j - x_i)$, donde $x_1 = 0, x_2 = t_{2,1} + s_{\psi(2),\psi(1)}, \dots, x_n = t_{2,1} + s_{\psi(2),\psi(1)} + \dots + t_{n,n-1} + s_{\psi(n),\psi(n-1)}$ y $t_{j,i} = t_j - t_i$, $s_{\psi(j),\psi(i)} = s_{\psi(j)} - s_{\psi(i)}$. La función $f_n(\theta_2) = g_n \theta_2^{n-1} [2 - \theta_2(t_{n,1} + s_{\psi(n),\psi(1)})]$ solamente depende de $s_{\psi(i)}$ con $i \in \{1, n\}$ y $g_n > 0$. De este modo, el determinante $|\Sigma_n|$ es siempre positivo si y sólo si $0 < \theta_2 < 2/(\lambda(\chi_1) + \lambda(\chi_2))$.

La demostración para el segundo caso es inmediata usando la regla de la correlación producto. \square

5.3.1. Función potencial modificada

En el marco de los diseños Marginalmente Restringidos trabajamos con dos variables explicativas (s, t) , siendo s la variable no controlable pero conocida y t la variable que esta bajo el control del experimentador. Consideremos la función:

$$q(s, t) = \left(\frac{1}{\theta_2} - 1 \right) |s|^{\theta_2} + |t|^{\theta_2}, \quad 0 < \theta_2 < 1.$$

Introducimos la covarianza potencial modificada:

$$\sigma^2 \exp[-q(s - s', t - t')].$$

El parámetro θ_2 es la medida de la influencia de la variable no controlable en la correlación. Si θ_2 es pequeño (próximo a 0), entonces hay una gran influencia de la variable no controlable. Por otro lado, si $\theta_2 \rightarrow 1$ la influencia es muy pequeña.

5.3.2. D -optimización y correlación

En la literatura, podemos encontrar diversas aplicaciones de diseño óptimo de experimentos con distintas funciones criterio para modelos espaciales de segundo orden. En este capítulo, consideramos el criterio de D -optimización que corresponde a la maximización del determinante de la matriz de información de Fisher.

Podemos encontrar justificaciones teóricas para usar la información de Fisher cuando las varianzas de los errores son pequeñas (ver Pázman [Paz04]). Además existen algunas justificaciones asintóticas, (Abt and Welch [Abt98]), donde se prueba tanto analíticamente como por simulación que la inversa de la matriz de información de Fisher es una buena aproximación del error cuadrático medio de la matriz de covarianzas de los estimadores de los parámetros en algunos casos especiales. Ellos suponen que $y(t)$ son observadas en $t_i = (i - 1)/(n - 1)$ para $i = 1, \dots, n$ en $\chi = [0, 1]$, pero sus resultados también se verifican si el conjunto de puntos $\{t_i\}$ es denso en χ cuando $n \rightarrow \infty$. Suponen una matriz de covarianza de la forma $\sigma^2 R$, donde $(R)_{i,j}$ puede corresponder a la estructura triangular, exponencial o Gaussiana.

Asimismo, Zhu y Stein [Zhu04] usan simulaciones para estudiar cuándo la inversa de la matriz de información de Fisher es una aproximación razonable de la matriz de covarianzas de los estimadores máximo verosímiles. Observaron que la matriz de información de Fisher da una buena estimación de la varianza de los

estimadores máximo verosímiles cuando el tamaño muestral es suficientemente grande.

Aunque algunas simulaciones y estudios teóricos muestran, que la inversa de la matriz de información de Fisher no es una buena aproximación de la matriz de covarianzas de los estimadores máximo verosímiles, puede usarse como criterio de diseño si la relación entre estos dos es monótona (véase Zhu y Stein [Zhu04]).

5.4. Regresión lineal con observaciones correladas

Consideremos el modelo $E[y(x)] = \alpha + \beta s + \gamma t$, donde $(s, t) \in \chi_1 \times \chi_2 = [0, 1] \times [0, 1]$. Suponiendo un diseño marginal $\xi_1 = \{0.3, 0.5, 0.7\}$, nuestro objetivo consiste en buscar diseños de 3-puntos $(0.3, t_1), (0.5, t_2), (0.7, t_3)$ maximizando la función criterio.

En primer lugar, consideremos la función de covarianza triangular 1.) del Lema 1 para diferentes valores nominales del parámetro $\theta_2 = 0.1, 0.2, \dots, 0.9$ ($\theta_2 \leq 1$ para que la covarianza sea no negativa) y la función de covarianza triangular separable 2.) con valores nominales del parámetro $\theta_2 = 0.1, 0.3, 0.5, 0.7, 1$ ($\theta_2 \leq 2$ para que la correlación sea no negativa). En ambos casos encontramos que el parámetro de covarianza tiene un diseño D-óptimo MR de la forma $t_1 = 0, t_2 = 1$ y $t_3 = 0$ tanto si los parámetros de la función de covarianza están fijos como si son de interés.

Por otro lado, si la función de correlación es la *Gaussiana* separable $\exp[-\theta_2(|s - s'|^2 + |t - t'|^2)]$, ($\theta_2 > 0$ para una correlación positiva), obtenemos los diseños dados en la Tabla 5.1. Esta puede considerarse un caso particular de la función de correlación potencial (5.1).

Si la función de correlación es la *exponencial* separable, $\exp[-\theta_2(|s - s'| + |t - t'|)]$ ($\theta_2 > 0$ para una correlación positiva), obtenemos los diseños de la Tabla 5.2. Estos diseños se ven afectados cuando el parámetro θ_2 sea o no considerado de interés. Observando los valores

j	θ_2	t_1	t_2	t_3	Inf_j
1	0.01	1	0	1	2529.086
2	0.01	1	0.5	0.2	5.272×10^7
1	0.1	0	1	0	28.023
2	0.1	0	0.8	0.1	5491.123
1	0.5	0	1	0	1.711
2	0.5	0	1	1	10.675
1	1	0	1	0	0.675
2	1	0	0	1	0.851
1	5	0	1	0	0.201
2	5	1	1	0	0.004
1	10	0	1	0	0.167
2	10	0	1	1	0.001
1	50	0	1	0	0.160
2	50	1	1	0	2.523×10^{-6}
1	100	0	1	0	0.160
2	100	1	1	0	4.299×10^{-8}

Tabla: 5.1: Diseños D-Óptimos MR cuando los parámetros de interés son los de la tendencia (primera fila para cada valor de θ_2 , $j = 1$) y cuando además de los parámetros del modelo, los de la función de covarianza son también de interés (segunda fila para cada valor de θ_2 , $j = 2$) y la correspondiente información.

j	θ_2	t_1	t_2	t_3	Inf_j
1	0.01	1	0	1	922.726
2	0.01	0	1	0	1.830×10^7
1	0.1	0	1	0	10.529
2	0.1	0	1	0	1941.076
1	0.5	0	1	0	0.723
2	0.5	0	1	0	4.017
1	1	0	1	0	0.326
2	1	0	1	0	0.334
1	5	0	1	0	0.163
2	5	0	1	0	0.001
1	10	0	1	0	0.160
2	10	1	0	0	6.309×10^{-5}
1	50	1	0	1	0.160
2	50	0	1	1	6.596×10^{-13}
1	100	0	1	0	0.160
2	100	1	1	0	1.359×10^{-20}

Tabla: 5.2: Diseños D –óptimos MR cuando los parámetros de interés son los del modelo (primera fila para cada valor de θ_2 , $j = 1$) y cuando ambos parámetros los del modelo y los de la función de covarianza son de interés (segunda fila para cada valor de θ_2 , $j = 2$) y la información correspondiente.

de la información en ambos casos, se observa que el parámetro θ_2 es muy sensible a valores extremos del parámetro.

Por otro lado, para valores del parámetro próximos a 1 y a 5 los diseños son idénticos. Para valores distintos de θ_2 , los diseños son distintos con algunos saltos notables en la información cuando θ_2 es de interés.

En los ejemplos, vemos que los diseños D –óptimos MR dependen del parámetro de correlación, θ_2 , cuando la correlación tiene una estructura Gaussiana y/o exponencial.

Además cabe destacar que hasta el momento, no hemos incluido

observaciones repetidas en el diseño marginal, sin embargo existen situaciones en las que es razonable permitir medidas repetidas. De modo que consideremos esta nueva situación para el modelo de regresión lineal. Debido a la posibilidad de falta de microvariación para las funciones de covarianza, introduciremos en ellas el llamado *efecto nugget* (ver Stehlík y otros [Ste05a]).

De modo que trataremos de encontrar diseños de 6 puntos $(0.3, t_1^{(1)})$, $(0.3, t_1^{(2)})$, $(0.5, t_2^{(1)})$, $(0.5, t_2^{(2)})$, $(0.7, t_3^{(1)})$, $(0.7, t_3^{(2)})$ con las siguientes estructuras de covarianza (con nugget effects):

1. Triangular 1.)

$$\text{cov}(y, y') = \begin{cases} \sigma^2(1 - \theta_2(|s - s'| + |t - t'|)), & \text{si } t \neq t' \text{ o } s \neq s'; \\ \sigma^2 + \alpha_s, & \text{si } t = t' \text{ y } s = s'. \end{cases}$$

2. Triangular 2.)

$$\text{cov}(y, y') = \begin{cases} \sigma^2(1 - \theta_2|s - s'|)(1 - \theta_2|t - t'|), & \text{si } t \neq t' \text{ o } s \neq s'; \\ \sigma^2 + \alpha_s, & \text{si } t = t' \text{ y } s = s'. \end{cases}$$

3. Gaussiana

$$\text{cov}(y, y') = \begin{cases} \sigma^2 \exp\{-\theta_2(|s - s'|^2 + |t - t'|^2)\}, & \text{si } t \neq t' \text{ o } s \neq s'; \\ \sigma^2 + \alpha_s, & \text{si } t = t' \text{ y } s = s'. \end{cases}$$

4. Exponencial

$$\text{cov}(y, y') = \begin{cases} \sigma^2 \exp\{-\theta_2(|s - s'| + |t - t'|)\}, & \text{si } t \neq t' \text{ o } s \neq s'; \\ \sigma^2 + \alpha_s, & \text{si } t = t' \text{ y } s = s'. \end{cases}$$

donde $\alpha_{0.3} = 0.1$, $\alpha_{0.5} = 0.5$, $\alpha_{0.7} = 1$.

Estas estructuras son coherentes, es decir, las matrices de covarianzas son definidas no negativas si y sólo si 1.) $0 < \theta_2 < 1$, 2.) $0 < \theta_2 < 2$, 3.) $0 < \theta_2$ y 4.) $0 < \theta_2$. Bajo este marco, los diseños que obtenemos aparecen en las Tablas 5.3-5.6. Una vez más, observamos cambios en los diseños, y saltos en la información. De nuevo, hay un rápido decrecimiento en la información con los valores del parámetro θ_2 cuando ambos parámetros son de interés. Este decrecimiento es más moderado para 1.).

El cálculo de estos diseños lo hemos realizado discretizando el espacio de diseño $\chi_2 = [0, 1]$ en intervalos de longitud 0.1, usando el software Matlab 7.

j	θ_2	$t_1^{(1)}$	$t_1^{(2)}$	$t_2^{(1)}$	$t_2^{(2)}$	$t_3^{(1)}$	$t_3^{(2)}$	Inf_j
1	0.01	0	1	1	0	1	0	2.083
2	0.01	0	1	1	0	1	0	286.392
1	0.1	0	1	0	1	0	1	0.915
2	0.1	0	1	0	1	1	0	32.762
1	0.5	1	0	0	1	0	1	0.228
2	0.5	1	0	0.5	0.5	0	1	1.226
1	1	1	0	1	0	1	0	0.154
2	1	1	0	1	0.4	0	1	0.715

Tabla: 5.3: Diseños D —óptimos MR para la estructura de covarianza 1.) cuando los parámetros de interés son los del modelo (primera fila para cada valor de θ_2 , $j = 1$) y cuando ambos parámetros los del modelo y los de la función de covarianza son de interés (segunda fila para cada valor de θ_2 , $j = 2$) y la información correspondiente.

j	θ_2	$t_1^{(1)}$	$t_1^{(2)}$	$t_2^{(1)}$	$t_2^{(2)}$	$t_3^{(1)}$	$t_3^{(2)}$	Inf_j
1	0.01	0	1	0	1	1	0	2.083
2	0.01	0	1	0	1	1	0	285.222
1	0.1	0	1	0	1	0	1	0.926
2	0.1	0	1	0	1	0	1	31.847
1	0.5	0	1	1	0	1	0	0.267
2	0.5	0	1	0.7	0	1	0	1.196
1	1	0	1	0	1	0	1	0.216
2	1	0	1	0	1	0	1	0.641

Tabla: 5.4: Diseños D —óptimos MR para la estructura de covarianza 2.) cuando los parámetros de interés son los del modelo (primera fila para cada valor de θ_2 , $j = 1$) y cuando ambos parámetros los del modelo y los de la función de covarianza son de interés (segunda fila para cada valor de θ_2 , $j = 2$) y la información correspondiente.

j	θ_2	$t_1^{(1)}$	$t_1^{(2)}$	$t_2^{(1)}$	$t_2^{(2)}$	$t_3^{(1)}$	$t_3^{(2)}$	Inf_j
1	0.01	0	1	1	0	0	1	2.109
2	0.01	0	1	1	0	0	1	275.488
1	0.1	0	1	1	0	1	0	1.048
2	0.1	0	1	1	0	1	0	29.444
1	0.5	1	0	1	0	1	0	0.383
2	0.5	0	1	0.5	0.5	1	0	0.764
1	1	0	1	0	1	1	0	0.268
2	1	1	0	0.5	0.5	0	1	0.169

Tabla: 5.5: Diseños D -óptimos MR para la estructura de covarianza Gaussiana cuando los parámetros de interés son los del modelo (primera fila para cada valor de θ_2 , $j = 1$) y cuando ambos parámetros los del modelo y los de la función de covarianza son de interés (segunda fila para cada valor de θ_2 , $j = 2$) y la información correspondiente.

j	θ_2	$t_1^{(1)}$	$t_1^{(2)}$	$t_2^{(1)}$	$t_2^{(2)}$	$t_3^{(1)}$	$t_3^{(2)}$	Inf_j
1	0.01	0	1	1	0	0	1	2.084
2	0.01	0	1	1	0	0	1	280.179
1	0.1	0	1	0	1	1	0	0.953
2	0.1	0	1	0	1	0	1	28.541
1	0.5	0	1	1	0	0	1	0.306
2	0.5	1	0	0.5	0.5	0	1	0.859
1	1	1	0	1	0	1	0	0.217
2	1	1	0	0.8	0.3	1	0	0.189

Tabla: 5.6: Diseños D -óptimos MR para la estructura de covarianza exponencial cuando los parámetros de interés son los del modelo (primera fila para cada valor de θ_2 , $j = 1$) y cuando ambos parámetros los del modelo y los de la función de covarianza son de interés (segunda fila para cada valor de θ_2 , $j = 2$) y la información correspondiente.

5.5. Retención radioactiva en el cuerpo humano

La función criterio más frecuente para modelos lineales y no lineales es la función de D -optimización. Este criterio maximiza el determinante de la matriz de información de los parámetros a estimar. Distintos ejemplos para modelos no lineales aparecen por ejemplo en Atkinson y Donev [Atk92]. En todos ellos se asumen errores independientes, sin embargo no siempre esta hipótesis es adecuada. Consideremos el problema de la asignación óptima de muestras para detectar el caso de una incorporación instantánea de sustancia radioactiva en los trabajadores de una fábrica. El modelo usado es una suma de muchas exponenciales (ver López-Fidalgo y otros [Lóp04c]), aunque sabemos que un número menor de exponenciales tienen un comportamiento similar. En nuestro ejemplo vamos a considerar una única exponencial para simplificar el problema. Las variables del modelo son el tamaño de la partícula, s , y el tiempo, t , en el que evaluamos la actividad biológica de la sustancia en el organismo. Claramente podemos suponer en este caso que las observaciones son correladas, puesto que son observaciones procedentes de un mismo trabajador en distintos tiempos. De modo que se pretende estimar la retención pulmonar de esa sustancia a tiempo t y tamaño s mediante el modelo:

$$\eta(\gamma, s, t) = \gamma e^{-\alpha s - \beta t}.$$

El tamaño de partícula s representa la Activity Median Aerodynamic Diameter (AMAD), que varía entre 1 y 10 unidades. En la práctica, el tiempo para la evaluación no debería pasar de los 100 días, y al menos debe pasar medio día desde que ocurrió el accidente para poder hacer la primera evaluación, de modo que podemos considerar como espacio de diseño el espacio producto, $\chi = [0, 10] \times [0.5, 100]$.

Supongamos que ocurre un accidente y que hay tres tamaños de partículas distribuidas en distintas zonas de la fábrica, de modo

que el diseño marginal es $\tilde{\xi}_1 = \{2, 5, 10\}$. Supongamos además que hay tres trabajadores, cada uno de los cuales está en una de las zonas anteriores y que podemos obtener dos muestras de cada uno de ellos en dos tiempos óptimos, como ya se apuntó más arriba. Las medidas que proceden del mismo trabajador se supone que serán correladas. Esto lo modelizaremos mediante una función de covarianza considerando un nugget effect, del siguiente modo:

$$\begin{aligned} \text{Cov}(y_{(t,s)}, y_{(t',s')}) &= \sigma^2(1 - \theta_2|t - t'|) && \text{si } s \neq s' \text{ o } t \neq t' \\ &= \sigma^2\alpha_s && \text{si } s = s' \text{ y } t = t'. \end{aligned}$$

Pueden obtenerse valores nominales para los parámetros de las estimaciones de los datos: $\gamma_0 = 0,1450$, $\alpha_0 = 0,1697$, $\beta_0 = 0,005783$. En la Tabla 5.7 podemos ver los diseños óptimos obtenidos para distintos valores de $\theta_2 > 0$, cuando los parámetros de interés sean los del modelo y cuando además de estos, sean los de la función de covarianza.

El cálculo de este diseño lo hemos realizado discretizando el espacio de diseño $\chi_2 = [0.5, 100]$ en intervalos de longitud 9.95.

5.6. Un procedimiento iterativo

En el contexto de las observaciones correladas, existen muy pocos resultados en relación al cálculo de diseños D -óptimos. Para obtener diseños óptimos exactos, proponemos en esta sección un procedimiento iterativo heurístico, en el sentido de que en cada paso mejoramos el diseño. Sin embargo no se tiene una demostración de la convergencia. Este algoritmo es una adaptación del presentado por Brimkulov, Krug y Savanov [Bri80] (en ruso). Una versión en inglés la encontramos en Uciniski y Atkinson, [Uci04]. Es un algoritmo de “intercambio” que empieza a partir de un diseño inicial $\xi^{(0)}$ y en cada iteración, un punto del soporte es reemplazado por otro del espacio de diseño de modo que incremente el determinante de la matriz de información. Una adaptación del algoritmo a nuestro caso, se podría resumir del siguiente modo:

j	θ_2	$t_1^{(1)}$	$t_1^{(2)}$	$t_2^{(1)}$	$t_2^{(2)}$	$t_3^{(1)}$	$t_3^{(2)}$	Inf_j
1	0.0001	0.5	100	0.5	0.5	0.5	0.5	14.211
2	0.0001	0.5	100	0.5	0.5	0.5	0.5	1.701×10^7
1	0.005	0.5	100	0.5	0.5	0.5	100	3.301
2	0.005	0.5	100	0.5	0.5	0.5	100	1.115×10^5
1	0.01	0.5	100	0.5	0.5	0.5	100	3.031
2	0.01	0.5	100	0.5	0.5	100	0.5	5.157×10^4
1	0.1	0.5	20.4	20.4	0.5	20.4	0.5	3.433
2	0.1	20.4	80.1	60.2	100	0.5	80.1	2.431×10^8
1	0.5	40.3	20.4	80.1	40.3	0.5	60.2	193.533
2	0.5	40.3	20.4	80.1	40.3	100	60.2	2.225×10^{14}
1	1	20.4	40.3	40.3	100	0.5	20.4	21.368
2	1	20.4	40.3	40.3	100	0.5	20.4	6.768×10^{10}

Tabla: 5.7: D -Óptimos MR cuando los parámetros de interés son los del modelo (primera fila para cada valor de θ_2 , $j = 1$) y cuando son de interés los parámetros del modelo y de la función de covarianza (segunda fila para cada valor de θ_2 , $j = 2$) y la información correspondiente.

Sea $\tilde{\xi}_1 = \{s_1, \dots, s_n\}$ un diseño marginal. Como vimos en el capítulo 2, el objetivo es encontrar los diseños condicionales $\xi_{2|1}(\cdot|s_i) = \{t_i\}$, $i = 1, \dots, n$ de modo que maximicen el determinante de la matriz de información del diseño conjunto. Supongamos que la estructura de covarianza es $C(|s_i - s_j|, |t_i - t_j|, \theta_2)$. El parámetro θ_2 lo supondremos conocido, es decir, no estamos interesados en estimarlo. Por tanto, la matriz de información, cuyo determinante debemos maximizar es, $M = X'\Sigma^{-1}X$. Los principales pasos del algoritmo son los siguientes:

1. Seleccionamos un diseño inicial, $\xi_{2|1}^{(0)}(\cdot|s_i) = t_i^0$, $i = 1, \dots, n$ con $|M| \neq 0$.
2. Para $I = \{1, \dots, n\}$ calculamos $(i^*, t^*) = \arg\max_{(i,t) \in I \times \chi_2} \Delta(i, t)$, donde

$$\Delta(i, t) = \frac{\{|M(\xi_{t_i \rightleftharpoons t}^{(l)})| - |M(\xi^{(l)})|\}}{|M(\xi^{(l)})|}$$

y $t_i \rightleftharpoons t$ significa que cambiamos el punto t_i por t .

3. Paramos el algoritmo cuando $\Delta^* \leq \delta$, siendo δ una tolerancia dada.

La demostración de la convergencia en los procedimientos iterativos en la teoría de diseños con observaciones incorreladas, se basaba en el teorema de equivalencia. Sin embargo, como se indicó al principio de este capítulo, en este contexto no tenemos un teorema de equivalencia debido a que las propiedades de convexidad fallan.

Los detalles de la implementación del algoritmo que encontramos en Uciński y Atkinson [Uci04], son adaptados para nuestro caso de diseños restringidos. (Véase el apéndice A.2.).

5.6.1. Un ejemplo real

Calcularemos diseños para un ejemplo de la vida real considerado por Stein [Ste05b]. Un modelo estadístico representa la velocidad media diaria del viento irlandés desde 1961 hasta 1978 en función de

s	10	14	20	25	41	64	69	72	85	89	104
t	14.0	7.0	25.0	3.5	30.0	20.8	1.0	27.8	7.0	25.0	14.0

Tabla: 5.8: Diseño óptimo conjunto para el modelo de viento irlandés

la localización de 11 estaciones meteorológicas en Irlanda. Usamos como diseño marginal las 11 situaciones distintas de las estaciones, obtenidas de Gneiting [Gne02]:

$$\tilde{\xi}_1 = \{10, 14, 20, 25, 41, 64, 69, 72, 85, 89, 104\}.$$

La estructura de covarianza que consideran Gneiting [Gne02] y Stein [Ste05b] es:

$$\begin{aligned} \text{cov}(y, y') &= \begin{cases} (0,901|t - t'|^{1,544} + 1)^{-1} & \text{if } s = s', \\ 0,968(0,901|t - t'|^{1,544} + 1)^{-1} \exp \left[-\frac{,00134|s - s'|}{(0,901|t - t'|^{1,544} + 1)^{\theta_2/2}} \right] & \text{if } s \neq s'. \end{cases} \end{aligned}$$

Una estimación del parámetro de la covarianza es $\hat{\theta}_2 = 0.61$ que será usada para los cálculos. La obtención de un diseño D -óptimo MR de modo directo en este ejemplo es computacionalmente muy difícil. De modo que usaremos el algoritmo anterior, con $n = 11$ y considerando $\delta = 10^{-6}$ para parar el algoritmo. Usando el diseño inicial dado anteriormente, obtenemos un determinante para la matriz de información de 1.5471×10^8 . En 8 iteraciones obtuvimos los siguientes valores para los determinantes de la matriz de información (divididos por 10^8): 2.1409, 2.5705, 2.8312, 3.0959, 3.1438, 3.1674, 3.1674, 3.1675.

En la Tabla 5.8 se observa el diseño óptimo conjunto cuyo determinante para la matriz de información es $|M| = 3.1675 \times 10^8$.

Nota: El algoritmo ha sido usado para diseños condicionales unipuntuales. Para diseños condicionales con k puntos todo es similar. Un grado de mayor complejidad aparece en el caso de diseños condi-

cionales no balanceados. En el caso particular de observaciones correladas solamente para los mimos valores de s ,

$$C(|s_i - s_j|, |t_i - t_j|, \theta_2) = \begin{cases} C(|t_i - t_j|, \theta_2) & \text{si } s_i = s_j \\ 0 & \text{si } s_i \neq s_j \end{cases}$$

la matriz de información es $M = X_1' \Sigma_1^{-1} X_1 + \dots + X_k' \Sigma_k^{-1} X_k$, donde X_i es la matriz del diseño para el mismo valor de $s = s_i$ y Σ_i es la caja correspondiente de la matriz de covarianzas de los errores.

5.7. Procedimiento de densificación

Retomemos el ejemplo propuesto en el capítulo 3. Decíamos que el uso de las barras de uranio es frecuente como combustible en centrales nucleares. Las barras consisten en un tubo que contiene pastillas hechas de uranio con una densidad específica. Una parte importante de este procedimiento consiste en la sinterización de las pastillas. Estas deben tener una densidad y una porosidad específicas. Además los poros tienen que retener la fisión de los gases durante el proceso de calentamiento en la central. Para este propósito se añaden unos aditivos a la mezcla inicial. Después de un tiempo en el horno los aditivos son quemados formando los poros en las barras.

Nuevamente, escribimos las variables que se consideraron:

1. “Densidad inicial” x_1 , cuyo rango va desde 94.9 % al 96.7 %. Esta variable no está sujeta al control del experimentador, pero sus valores son conocidos antes de que la prueba sea realizada. Consideramos aquí una muestra de 392 medidas obtenidas de la fábrica de ENUSA.
2. “Porcentaje de aditivo U_3O_8 ”, que varía desde 0 a 20, en porcentajes estándar. Esta variable está completamente bajo el control del experimentador.

La variable respuesta considerada fue la densidad final. El modelo utilizado fue un modelo de regresión cuadrática para las dos variables explicativas, $y = \gamma_0 + \gamma_1 s + \gamma_2 t + \gamma_{12} st + \gamma_{11} s^2 + \gamma_{22} t^2$ y la estructura de la covarianza fue la potencial modificada:

$$\text{cov}(y, y') = \sigma^2 \exp[(1/\theta_2 - 1)|s - s'|^{\theta_2} + |t - t'|^{\theta_2}], \quad 0 < \theta_2 < 1.$$

Para obtener una estimación del parámetro θ_2 a partir de los datos, procedemos de la siguiente forma. Si se usan todos los datos, el determinante de la matriz de covarianzas estará próximo a cero. Para evitar problemas con el cálculo, consideramos una submuestra como describimos a continuación. Dividimos la muestra en clases con t constante. De este modo, obtenemos 18 clases tales que el número de elementos en las clases son $n_1 = 78$, $n_2 = n_{18} = 67$ y $n_i = 12$ en el resto de los casos. Entonces, representamos proporcionalmente las clases mediante las siguiente aproximaciones de submuestras representativas $m_1 = 6$, $m_2 = m_{18} = 5$ y $m_i = 1$, en el resto. Para $m_i = 1$ tomamos directamente la media aritmética de los valores de la clase que representa.

Finalmente maximizamos el logaritmo de la función de verosimilitud $L(\theta_2, y)$, usando la función de tendencia:

$$-2079,25 + 44,84s - 2,095t + 0,01st - 0,23s^2 + 0,00005t^2,$$

donde los coeficientes son estimados suponiendo las observaciones i.i.d. Después de una optimización numérica obtenemos que el estimador máximo-verosímil del parámetro θ_2 es $\hat{\theta}_2 = 0.93$.

De nuevo, la obtención de un diseño D-óptimo MR es complicada desde un punto de vista computacional. Como en la sección anterior, aplicamos una versión del algoritmo de Brinkulov para resolver el problema. El diseño óptimo conjunto fue calculado y el determinante de la matriz de información del diseño óptimo resultó ser $|M| = 5.6834 \times 10^{14}$. (El diseño conjunto aparece en el apéndice A.3.)

$\theta_2^{(G)} \backslash \theta_2^{(E)}$	0.01	0.1	0.5	1.0	5.0
0.01	1	1	0.881	0.881	0.896
0.1	1	1	0.962	0.962	0.961
0.5	0.965	0.965	1	1	0.977
1.0	0.962	0.962	0.975	0.975	0.999
5.0	1	1	0.639	0.639	0.823

Tabla: 5.9: D-eficiencias de diseños D -óptimos bajo la estructura de covarianza Gaussiana con valor nominal del parámetro $\theta_2^{(G)}$, cuando la verdadera estructura de covarianza es la exponencial con valor nominal del parámetro $\theta_2^{(E)}$.

5.8. Análisis de sensibilidad

En esta sección estudiamos la sensibilidad de los diseños D -óptimos ante la selección de los valores nominales del parámetro θ_2 y ante el cambio de la estructura de covarianza.

En particular, consideramos los diseños para el modelo lineal (Sección 5.4). En la Tabla 5.9 se presentan las D-eficiencias de los diseños D -óptimos calculados, suponiendo la estructura de covarianza Gaussiana para distintos valores nominales de $\theta_2^{(G)}$ cuando la verdadera estructura de covarianza es la exponencial con valores nominales del parámetro $\theta_2^{(E)}$.

Obtenemos una eficiencia de 1 para todos los casos cuando los parámetros de interés son solamente los de la tendencia, sin embargo, cuando los parámetros de interés son los de la tendencia y los de la covarianza, entonces las eficiencias aparecen en la Tabla 5.9. Estas son altas, excepto para los valores extremos $\theta_2^{(G)} = 5.0$ y $\theta_2^{(E)} = 0.5$ o 1 en una posición central.

A continuación estudiamos los cambios en la eficiencia cuando el valor nominal de θ_2 no es preciso. En la Tabla 5.10 aparecen las D-eficiencias de diseños D -óptimos suponiendo que el valor del parámetro es θ_2 cuando el verdadero valor es θ_2^* con la misma estructura de covarianza (exponencial y Gaussiana respectivamente)

y cuando ambos parámetros los de la tendencia y de la covarianza son de interés.

Por último la Tabla 5.11 muestra las eficiencias para valores erróneos de los valores nominales en el ejemplo de la retención radioactiva del ejemplo de la sección 5.5. Todas ellas son muy altas, de modo que podríamos concluir diciendo que no deberíamos preocuparnos por las elecciones erróneas del parámetro θ_2 .

$\theta_2 \backslash \theta_2^*$	0.01		0.1		0.5		1.0		5.0	
	E	G	E	G	E	G	E	G	E	G
0.01	1	1	1	1	0.881	0.863	0.908	0.863	1	0.885
0.1	1	1	1	1	0.962	0.928	0.965	0.928	1	0.941
0.5	0.965	0.992	0.965	0.992	1	1	0.995	1	0.966	0.993
1.0	0.962	0.902	0.962	0.902	0.975	1	1	1	0.962	0.958
5.0	1	0.787	1	0.787	0.642	0.919	0.845	0.919	1	1

Tabla: 5.10: D-eficiencias bajo la suposición de que el valor del parámetro es θ_1^* para las estructuras de covarianza exponencial (E) y Gaussiana (G) cuando ambos parámetros el de la tendencia y el de la covarianza son de interés

$\theta_2 \backslash \theta_2^*$	0.0001		0.005		0.01	
j	1	2	1	2	1	2
0.0001	1	1	0.982	0.999	0.982	0.992
0.005	0.960	0.958	1	1	1	1
0.01	0.886	0.910	1	1	1	1

Tabla: 5.11: D-eficiencias para el ejemplo de la Sección 5.5 cuando solamente el parámetro de la tendencia es de interés ($j = 1$) y cuando ambos el de la tendencia y el de la covarianza son de interés ($j = 2$) cuando el verdadero valor del parámetro es θ_2^* y el que fue usado para calcular el diseño óptimo fue θ_2

5.9. Conclusiones

En este capítulo hemos considerado el problema de diseñar cuando las observaciones son correladas y una de las variables explicativas no está bajo el control del experimentador. Un problema no sencillo, consiste en encontrar una estructura de covarianza apropiada para un caso particular. Un estudio de distintas estructuras de covarianza ha sido realizado para ayudar al experimentador a tomar una decisión. Asimismo hemos realizado un análisis de sensibilidad frente a selecciones erróneas de la estructura de la covarianza y valores nominales de los parámetros.

Considerando el modelo de regresión lineal hemos estudiado diferentes estructuras de covarianza: Dos tipos de estructuras triangulares, una Gaussiana y una exponencial. También hemos empleado la función de covarianza triangular utilizada en un problema real de retención radioactiva. La estructura que aparece en Gneiting [Gne02] y Stein [Ste05b] ha sido también considerada para el cálculo de diseños MR. Por último, hemos considerado una versión modificada de la estructura potencial, usada para la modelización de un procedimiento de densificación en la construcción de combustible para una planta nuclear. En todos los casos, los resultados son coherentes, pero difieren sustancialmente del caso en el que las observaciones son incorreladas, de este modo se muestra la importancia de usar una estructura de covarianza apropiada. Del mismo modo, existen diferencias significativas entre los diseños obtenidos cuando θ_1 era de interés y cuando los dos parámetros θ_1 y θ_2 tenían que ser estimados.

Para evitar que los diseños colapsen en un punto se introdujo el llamado efecto nugget.

Finalmente, para el cálculo de diseños D -óptimos MR con observaciones correladas se ha adaptado un algoritmo iterativo que proporciona diseños casi óptimos. Sin embargo, este algoritmo debe ser revisado para mejorar la rapidez en la convergencia.

Capítulo 6

Conclusiones e investigación futura

6.1. Conclusiones

En esta memoria, hemos considerado la construcción de diseños óptimos para varios ejemplos reales que han ido añadiendo distintas complicaciones a la teoría clásica de diseño óptimo. En especial, los diseños en los que aparecen una variable controlable y otra no controlable con valores conocidos antes de que el experimento sea realizado. Son los llamados diseños marginalmente restringidos, MR, que han sido estudiados en diferentes contextos (capítulos 4 y 5). La mayor parte de estos resultados, se han presentado en congresos internacionales y nacionales, y se encuentran publicados en los *proceedings* de los mismos. (Véase Ballesteros et al. [Bal05], López Fidalgo and Martín Martín [Lóp05] and [Lóp04b].)

Aunque en cada capítulo hemos tratado de realizar conclusiones a la finalización de los mimos, es útil hacer un resumen de todo el trabajo conjunto. A continuación, recordamos los principales objetivos de la investigación así como las contribuciones originales que hemos aportado:

1. En el **capítulo 1** se proporciona una introducción a la teoría de diseño de experimentos: propiedades de la matriz de información, distintos criterios y, por supuesto, los teoremas de equivalencia de Kiefer y Whittle. Como consecuencia de estos teoremas se consideran condiciones de optimalidad basadas en las derivadas direccionales. Por último se comparan las cotas para la D - eficiencia propuestas por Kiefer y Atwood.
2. Recopilación de los trabajos existentes en la literatura sobre el diseño de experimentos cuando algunas variables del modelo no están bajo el control del experimentador. En el **capítulo 2** hemos repasado la teoría de los diseños marginalmente restringidos, condicionalmente restringidos así como marginal y condicionalmente restringidos, resaltando los teoremas de equivalencia para cada uno de los casos, y los algoritmos propuestos para el cálculo de los mismos, pues en los problemas reales, el cálculo “directo” de los diseños óptimos podría resultar muy laborioso. Por otro lado, las aportaciones en este capítulo han sido: el análisis de las distintas funciones criterio para cada uno de los casos, indicando las distintas cotas para la eficiencia y la aportación de “nuevas” cotas para la D - eficiencia de los diseños MR, CR y MCR, basadas en la que encontramos en Atwood [Atw69].
3. En el **capítulo 3**, se han proporcionado una serie de métodos, basados en los algoritmos multiplicativos propuestos por Torsney [Tor01],[Tor77] $p_j^{(r+1)} \propto p_t^{(r)} f(d_j)$, siendo $d_j = \frac{\partial \Phi}{\partial p_j}$ (véase (3.1) y (3.2)), para determinar diseños óptimos marginalmente restringidos, así como los diseños marginal y condicionalmente restringidos.

Aunque no tenemos una prueba de la convergencia de estos algoritmos, estudios computacionales muestran una rápida convergencia, pero cerca del óptimo, la tasa de convergencia disminuye considerablemente (ver Tablas 3.1, 3.3, 3.7 y la sección final de 3.5.1).

Aunque la monotonicidad no garantiza la rapidez de conver-

gencia, una elección apropiada de la función $f(d)$ proporciona diseños casi óptimos en pocas iteraciones. Esto lo podemos observar en la Tabla 3.1 donde vemos que para diseños $D-$, D_A- , $c-$ y L_A- óptimos se obtiene $F_{2j|1i} \geq -0.1$ en una o dos iteraciones.

También podemos encontrar en este capítulo los Teoremas 3.1 y 3.2, que muestran la equivalencia entre diseños $D-$ óptimos MR o MCR y los diseños D_s- óptimos. Podemos ver en el Ejemplo 1 de la sección 3.3.2 que los diseños $c-$ óptimos y $D_{\theta_2}-$ óptimos son los mismos, ratificando de este modo el Teorema 3.1. Análogamente el Teorema 3.1 queda ilustrado por el ejemplo de la sección 3.4.2 donde el diseño $D-$ óptimo MCR para todos los parámetros, es también el $D_{\theta_2}-$ óptimo MCR cuando los parámetros de interés son los de la variable no controlable con valores desconocidos y el parámetro de la variable controlable.

Por último distintas aplicaciones al mundo de la industria y de la medicina se han presentado para el cálculo de los diseños usando los algoritmos que se han desarrollado.

4. En el **capítulo 4** se proporcionan algunos métodos para calcular diseños $D-$ óptimos marginalmente restringidos cuando el diseño marginal sigue una distribución discreta o continua. Los resultados de las secciones 4.2 y 4.3 pueden ser fácilmente generalizados cuando aparezca una nueva variable no controlable.

Se propone un algoritmo secuencial para el cálculo de diseños $D-$ óptimos MR. Suponiendo que hemos calculado nuestro diseño, pero que una nueva observación aparece, el diseño marginal cambia. Nuestro objetivo consiste en asignar un valor a la variable controlable de forma que el diseño conjunto sea óptimo. Los diseños secuenciales se construyen de tal modo que toda la información que se tiene en las etapas previas, es usada para sugerir cómo calcular la siguiente iteración.

5. En el **capítulo 5** estudiamos el problema de diseñar bajo la

posibilidad de que las observaciones sean correladas y una de las variables explicativas no está bajo el control del experimentador. Realizamos un estudio sobre distintas estructuras de covarianza para ayudar al experimentador a tomar una decisión. Asimismo hemos hecho un análisis de sensibilidad frente a selecciones erróneas de la estructura de la covarianza y valores nominales de los parámetros.

Considerando el modelo de regresión lineal hemos estudiado diferentes estructuras de covarianza: Dos tipos de estructuras triangulares, una Gaussiana y una exponencial. También hemos empleado la función de covarianza triangular utilizada en un problema real de retención de radioactividad. La estructura que aparece en Gneiting [Gne02] y Stein [Ste05b] ha sido también considerada para el cálculo de diseños MR. Por último, hemos empleado una versión modificada de la estructura potencial, usada para la modelización de un procedimiento de densificación en la construcción de combustible para una planta nuclear. En todos los casos, los resultados son coherentes, pero difieren sustancialmente del caso en el que las observaciones son incorreladas, de este modo se muestra la importancia de usar una estructura de covarianza apropiada. Del mismo modo, existen diferencias significativas entre los diseños obtenidos cuando θ_1 era de interés y cuando los dos parámetros θ_1 y θ_2 tenían que ser estimados.

Para evitar que los diseños se colapsen en un punto se introdujo el llamado efecto nugget.

Finalmente, para el cálculo de diseños D —óptimos MR con observaciones correladas se ha adaptado un algoritmo iterativo que proporciona diseños casi óptimos. Sin embargo, este algoritmo debe ser revisado para mejorar la rapidez en la convergencia. El código fuente realizado en *Matlab* se puede consultar en el apéndice A.2.

6.2. Líneas de investigación futuras

6.2.1. Modelos Lineales Generalizados

Los modelos lineales han sido ampliamente tratados en la teoría de diseño óptimo, entre otros motivos, debido a que la matriz de información para estos modelos es relativamente sencilla y no depende de los parámetros a estimar. Podemos pensar en los diseños Lineales Generalizados como una extensión del caso lineal. Por este motivo, además de las aplicaciones de los diseños no lineales, en la mayoría de las situaciones reales, se han estudiado los diseños lineales generalizados. Así Ford, Torsney y Wu [For92] construyen diseños c - y D -óptimos, usando propiedades geométricas, para modelos lineales generalizados con una variable de diseño. Sitter y Wu [Sit93] calculan diseños D - y A -óptimos para experimentos con respuesta binaria y con una variable de diseño. Sitter y Torsney, [Sit95], tratan una situación más general considerando varias variables en el diseño.

Aunque no ha sido un objetivo de esta tesis el cálculo de diseños para Modelos Lineales Generalizados, introducimos una serie de conceptos para enmarcar una posible línea de investigación futura.

Sea Y una variable aleatoria. Se dice que su distribución pertenece a la *familia exponencial* si podemos escribir:

$$f(y; \theta) = s(y)t(\theta) \exp\{a(y)b(\theta)\},$$

siendo a, b, s y t funciones conocidas. Podemos escribir la ecuación anterior como:

$$f(y; \theta) = \exp\{a(y)b(\theta) + c(\theta) + d(y)\}.$$

Si $a(y) = y$ diremos que la distribución anterior está escrita en la forma canónica. Muchas de las distribuciones más conocidas, pertenecen a la familia exponencial (Poisson, Normal, Binomial...).

Consideremos ahora Y_1, \dots, Y_m , variables aleatorias independientes, en donde cada una de ellas tiene una distribución que pertenece a la

familia exponencial. Se verifica que la distribución de probabilidad conjunta es:

$$f(y_1, \dots, y_m; \theta_1, \dots, \theta_m) = \exp\left[\sum_{i=1}^m y_i b(\theta_i) + \sum_{i=1}^m c(\theta_i) + \sum_{i=1}^m d(y_i)\right].$$

Los modelos lineales generalizados, verifican que una combinación de los parámetros, es una función del valor esperado $E[Y_i] = \mu_i$, es decir,

$$\eta(\mu_i) = \theta^t x_i,$$

siendo x_i un vector de variables explicativas, θ el vector de parámetros y η una función monótona y diferenciable, denominada *función link*.

La matriz de información, viene dada por

$$I_{i,j} = -E\left[\frac{\partial^2 L(\theta; y)}{\partial \theta_i \partial \theta_j}\right]$$

siendo

$$L(\theta; y) = \sum y_i b(\theta_i) + \sum c(\theta_i) + \sum d(y_i),$$

$$E[Y_i] = \mu_i = \frac{-c'(\theta_i)}{b'(\theta_i)}$$

y

$$Var(Y_i) = \frac{b''(\theta_i)c'(\theta_i) - c''(\theta_i)b'(\theta_i)}{[b'(\theta_i)]^3}.$$

Por tanto, se puede probar que

$$I_{i,j} = \sum_{l=1}^m x_{li} x_{lj} w(\mu_i),$$

donde

$$w(\mu_i) = \frac{\left(\frac{\partial \eta}{\partial \mu_i}\right)^2}{var(Y_i)}.$$

Para un mayor desarrollo de esta teoría véase Dobson [Dob83].

Seguendo la notación que aparece en Ford, Torsney y Wu [For92], el problema que consideramos es el siguiente:

$$E[y(x)] = \eta(\theta, x),$$

siendo η una función no lineal en los parámetros de una función lineal en θ , es decir,

$$\eta = \eta(\theta^t s),$$

donde $s^t = (1, x^t)$. Por tanto, si χ es nuestro espacio de diseño, y ξ es un diseño aproximado, la matriz de información será

$$M(\xi) = \int_{\chi} w(\theta^t s) s s^t \xi(dx),$$

siendo

$$w(\theta^t s) = \frac{\left(\frac{\partial \eta}{\partial(\theta^t s)}\right)^2}{\text{var}(y \mid x)}.$$

En particular, consideremos el caso

$$\eta = \eta(\theta_1 + \theta_2 x), \quad \theta = (\theta_1, \theta_2) \quad s = (1, x)^t.$$

Realizando la transformación $s \rightarrow t = Bs$, siendo B una matrix no singular 2×2 , el espacio de diseño χ se transforma en otro espacio Z . De este modo transformamos el problema de diseño no lineal en uno lineal, es decir, el problema puede ser resuelto independientemente de θ . Así, resolviendo nuestro problema para cualquier Z , quedará resuelto para χ y cualquier θ . Aunque debemos resaltar que los únicos criterios que podemos utilizar son los que sean invariantes bajo la transformación anterior (como D - y c - optimización). Elegimos B tal que $t = (1, z)^t$, siendo

$$s \longrightarrow t = Bs = \begin{pmatrix} 1 & 0 \\ \theta_1 & \theta_2 \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix} = \begin{pmatrix} 1 \\ \theta_1 + \theta_2 x \end{pmatrix} = \begin{pmatrix} 1 \\ z \end{pmatrix}$$

Ford, Torsney y Wu [For92] utilizan criterios geométricos para la construcción de los diseños, basándose en el resultado de Silvey

[Sil80] que establece que el soporte de los diseños óptimos depende de la geometría del espacio definido por:

$$G = \{g \in \mathbb{R}^m : g = \sqrt{w(t_m)}t, t = (1, z^t)^t, z \in Z\},$$

en nuestro caso

$$G = \{g(z) = (g_1, g_2)^t / g_1 = \sqrt{w(z)}, g_2 = zg_1, z \in Z\}.$$

Los pasos a seguir para encontrar los diseños D -óptimos son los siguientes:

1. Tomar una decisión sobre el número de puntos del soporte.
2. Maximizar el determinante de la matriz de información conocido el número de puntos del diseño óptimo.

Según encontramos en Silvey [Sil80], los puntos del soporte del óptimo, son los puntos de contacto entre el conjunto G y el menor elipsoide centrado en el origen que contiene a G .

Considerando el modelo

$$E[y(x_1, x_2)] = \eta(\theta_1 x_1 + \theta_2 x_2),$$

donde x_1 es una variable no controlable, pero con valores conocidos antes de realizar el experimento. Nos planteamos la aplicación de la teoría de los modelos lineales generalizados al estudio del cálculo de diseños óptimos MR.

Realizando el cambio:

$$s \longrightarrow t = Bs = \begin{pmatrix} \theta_1 & \theta_2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \theta_1 x_1 + \theta_2 x_2 \\ x_1 \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix},$$

lo que buscamos como vimos en el capítulo 2 son los diseños condicionales $\xi_{2|1}$, para cada x_1 (y por tanto z_2) fijo.

En este caso:

$$G = \{g(z) = (g_1, g_2)^t / g_1 = \sqrt{w(z_1)}z_1, g_2 = \sqrt{w(z_1)}z_2, z \in Z = Z_1 \times Z_2\}.$$

Si encontramos valores óptimos z_1^* dado z_2 , podemos encontrar $x_2^* = \frac{z_1^* - \theta_1 x_1}{\theta_2}$.

Algunos problemas abiertos serían los siguientes:

1. Caracterización geométrica de los diseños condicionales D –óptimos.
2. Cálculo de diseños para modelos lineales generalizados cuando una de las variables en el modelo sea no controlable.

6.2.2. Diseños con observaciones correladas

En el capítulo 5 hemos calculado diseños D –óptimos MR cuando las observaciones son correladas con una función de varianza dada. En algunos casos, vimos que para el cálculo de estos diseños eran necesarios algoritmos debido a la complejidad de los cálculos. Uno de ellos fue el propuesto por Brimkulov y otros [Bri80]. Desafortunadamente, la convergencia de este algoritmo al diseño óptimo no está garantizada. Müller y Pázman [Mü99], proponen un algoritmo distinto para el cálculo de estos diseños, basado en una nueva interpretación de las medidas de diseño propuesta por ambos autores [Paz98]. Este algoritmo es también un algoritmo de intercambio de un punto, y está basado en la derivada direccional. La única restricción con esta nueva interpretación de las medidas de diseño es que no se permiten réplicas. Un nuevo problema consistiría en adaptar este tipo de algoritmos al caso de los diseños MR, y hacer una comparación entre ambos algoritmos.

Cuando las observaciones son incorreladas, una forma para “medir” la bondad de un diseño es mediante su eficiencia, y como ya dijimos, para poder calcularla necesitamos conocer quién es el diseño

óptimo. Por este motivo, introducíamos cotas inferiores para la eficiencia. Bajo el marco de las observaciones correladas, no se conoce ninguna medida de la cota de la eficiencia para un diseño. Solamente Bischoff [Bis95] proporciona una cota pero cuando las matrices de covarianza son simétricas regulares circulares (es decir cada fila se obtiene desplazando cíclicamente una posición la fila anterior). Una línea de investigación abierta sería buscar alguna medida de esta cota para casos más generales.

Aunque en las secciones anteriores hemos planteado las líneas de investigación más relevantes, otras posibles cuestiones serían:

- Aplicación de la teoría bayesiana al cálculo de diseños MR.
- Elaboración de nuevas teorías cuando nuevas restricciones aparezcan en el modelo desarrollando nuevos algoritmos para su resolución.
- Búsqueda de condiciones suficientes para garantizar la convergencia del algoritmo secuencial propuesto en el capítulo 4.

Conclusions and Future Work

Conclusions

In this thesis, the construction of optimal designs with constraints on the variables of the model, has been considered. Through real examples, different situations have been presented. One of the most important appears when in the model there are two variables: one variable is controlled and the other is uncontrolled but its values are known before the experiment is carried out. These designs are called marginally restricted designs(MR). Not only have we computed MR optimal designs (Chapter 3) but we have also studied these designs under several frameworks (sequentially designs and designs for correlated observations in Chapter 4 and 5, respectively).

Most of this work has been presented at international and national congress and they have been published in the proceeding of them. (See [Bal05], [Lóp05] y [Lóp04b]).

The principal targets and contributions of this thesis are summarized in the following:

1. An introduction to the theory of optimal design (properties of the information matrix, several criteria and the equivalence theorems) is provided in Chapter 1. Also we discussed optimality conditions based on directional derivatives, and two lower bound for the efficiency are considered. One of them

is proposed by Kiefer [Kie59a] and the other one is proposed by Atwood [Atw69]. We proved that this last bound is better than Kiefer's bound.

2. A review of the works on optimal designs with constraints is made in Chapter 2. The properties of optimum experimental designs when the information matrix depends upon uncontrolled variables are described. Three cases are differentiated (marginally restricted (MR), conditionally restricted (CR) and marginally conditionally restricted (MCR) designs) and their equivalence theorems are given. We go into detail about different criteria. We highlight the algorithms derived from equivalence theorems presented in López Fidalgo and Garcet Rodríguez [Lóp04a]. Finally, the bound for the efficiency proposed by Atwood is generalized to MR, CR and MCR. This bound is higher than the bound proposed by López Fidalgo and Garcet Rodríguez [Lóp04a]. Examples are given.
3. The purpose of Chapter 3 is to address the question of constructing optimal designs when some of the factors are known in advance. We have proposed methods for computing such designs, in particular, a class of multiplicative algorithms is considered in Chapter 3 in order to compute MR and MCR optimal designs. These algorithms are based on the iteration proposed by Torsney [Tor77] $p_j^{(r+1)} \propto p_t^{(r)} f(d_j)$, where $d_j = \frac{\partial \Phi}{\partial p_j}$ (see (3.1) and (3.2)). Although monotonicity does not guarantee fast convergence, an appropriate choice of function $f(d)$ produces almost optimal designs in a few iterations. Finally, it is important to mention a related issue that deserves further attention. Theorems 3.1. and 3.2. show the equivalence between D –optimal MR or MCR designs and the appropriate D_s –optimality.
4. In Chapter 4 MR D –optimal designs have been provided. The results in Sections 4.2 and 4.3 can be easily generalized when a non controlled new variable is observed. The sequen-

tial algorithm proposed in this paper was motivated by the example which appears in López-Fidalgo and Garcet-Rodríguez [Lóp04a]. They proposed an optimal experimental design for predicting morbidity after lung resection. The main variables considered in this paper were the “percentage of maximum volume of expired air”, the “oxygen desaturation during the test” and the “exercise time in minutes”. We consider a more simplified problem only taking into account the first and the third variables. Assuming that we have calculated our design, but a new patient arrives, then the marginal design changes. We want to assign an exercise time in such a way that the joint design is optimal. The sequential designs are constructed in such a way that all of the information in the previous runs can be used efficiently in suggesting how the next run should be conducted.

5. In Chapter 5 we have considered the problem with correlated observations when one of the explanatory variables is not under the control of the experimenter (MR designs). A nontrivial problem is to find the appropriate covariance structure for a particular case. A study of covariance structures has been made in order to help the experimenter make the decision. A sensitive analysis against bad choices of the covariance structure and the nominal values of the parameters was performed. Considering linear regression for the mean, four different covariance structures have been studied: Two types of triangular structures, a Gaussian and an exponential structure. A triangular structure was also used in a real radiation retention problem. The structure considered by Gneiting [Gne02] and Stein [Ste05b] was also investigated for MR designs. Finally, a modified version of the powered structure has been used for a densification procedure in the construction of fuel for nuclear plants. In all cases the results are coherent, but differ from the results for the uncorrelated case, showing the importance of using an appropriate covariance structure. In the same way there are significant differences between the designs obtained

when θ_1 (parameters of the trend) was of interest and when both θ_1 and θ_2 (parameters of the covariance structure) were to be estimated. They are also coherent with the experience of the practitioners for the real problems considered. A nugget effect was needed to prevent the designs from collapsing at one point. An algorithm was adapted to this situation and proved to be efficient. Nevertheless some more work on this may be done to improve the rates of convergence.

Future Work

Generalized Linear Models

Reviewing the literature on the optimal designs of experiments, we can see that the majority of the researches on this topic remains focussed on linear models. The literature on optimal design for non linear models has been relatively sparse due to the dependence on parameter values. There have been several papers on optimal design for the efficient estimation of parameters in generalized linear models (GLMs), but the literature in this area tends to be even more sparse than in the area of nonlinear models, due to the increased theoretical and computational challenges associated with GLMs. GLMs are commonly used in experiments in such areas as chemistry, pharmacology and engineering, so optimal design for these models warrants further consideration. The work of Ford, Torsney and Wu [For92] is of special interest, since they used a canonical form to express design problems and were able to find D - and c -optimal designs for one predictor. Sitter and Wu [Sit93] compute D - and A -optimal designs for experiments with binary response and one predictor. Sitter and Torsney [Sit95] compute D -optimal designs for multiple predictors.

Even though the computation of designs for GLMs is not the focus of this thesis, we are going to introduce a series of concepts (see Ford, Wu and Torsney [For92]) that may form the bases of future lines of investigations.

GLMs encompass a large class of models, from simple linear regression models to models for quantal responses to models for survival data. They can be studied as a single class.

In experimental design with non-linear models the scalar response variable y has distribution belonging to the exponential family:

$$E[y(x)] = \eta(\theta^t s).$$

η is a non-linear function of a linear function of θ , and $s^t = (1, x^t)$. Thus, if χ is the design space and ξ is a measure, the Fisher information matrix for this design will be:

$$M(\xi) = \int_{\chi} w(\theta^t s) s s^t \xi(dx),$$

where

$$w(\theta^t s) = \left(\frac{\partial \eta}{\partial \theta^t s} \right)^2 \bigg/ \text{var}(y | x).$$

In particular, we consider

$$\eta = \eta(\theta_1 + \theta_2 x), \quad \theta = (\theta_1, \theta_2) \quad s = (1, x)^t.$$

Carrying out the transformation $s \rightarrow t = Bs$, where B is a nonsingular 2×2 matrix, the design space χ is transformed into a design space Z . This way, we convert our non-linear problem in a linear problem. Then the problem can be solved independently of θ . Note that the only criteria that we can use are those that are invariant under the previous transformation (such as D - and c -optimality). B is chosen such as $t = (1, z)^t$, where

$$s \longrightarrow t = Bs = \begin{pmatrix} 1 & 0 \\ \theta_1 & \theta_2 \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix} = \begin{pmatrix} 1 \\ \theta_1 + \theta_2 x \end{pmatrix} = \begin{pmatrix} 1 \\ z \end{pmatrix}$$

Ford, Torsney and Wu used geometric criteria to construct designs. They are based on the result of Sylvey (1980) who established that the support of the optimal designs depends on the geometry of the space defined by:

$$G = \{g \in \mathbb{R}^m : g = \sqrt{w(t_m)}t, t = (1, z^t)^t, z \in Z\},$$

in our case

$$G = \{g(z) = (g_1, g_2)^t / g_1 = \sqrt{w(z)}, g_2 = zg_1, z \in Z\}.$$

The steps for finding D -optimal designs are the following:

1. Make a decision about the number of the support points.
2. Maximize the determinant of the Fisher information matrix.

Sylvey proposes that the support points are those that are in contact with the set G and the minimum ellipsoid centering on the origin containing G .

We consider the model

$$E[y(x_1, x_2)] = \eta(\theta_1 x_1 + \theta_2 x_2),$$

where x_1 is an uncontrolled variable whose values are known before the experiment is carried out and x_2 is a controlled variable. We can adopt the previous ideas to compute MR D -optimal designs:

$$s \longrightarrow t = Bs = \begin{pmatrix} \theta_1 & \theta_2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \theta_1 x_1 + \theta_2 x_2 \\ x_1 \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}.$$

For each x_1 (and then for each z_2) fixed, we are looking for conditional designs $\xi_{2|1}$, which are performed in practice.

Under the above transformation:

$$G = \{g(z) = (g_1, g_2)^t / g_1 = \sqrt{w(z_1)}z_1, g_2 = \sqrt{w(z_1)}z_2, z \in Z = Z_1 \times Z_2\}.$$

If we find optimal z_1^* values given z_2 , we can find $x_2^* = \frac{z_1^* - \theta_1 x_1}{\theta_2}$.

This raises the following issues:

1. Geometrical characterizations of D –optimal conditional designs.
2. Computation of optimal design for GLMs when some of the variables in the model are not under control of the experimenter.

Optimal Designs for correlated observations

Marginally restricted D –optimal designs for correlated observations have been calculated in Chapter 5, when a covariance function is given. In some cases directly obtaining a MR D –optimal design could be computationally impossible. There exist several algorithms for the calculation of optimum designs in the regression case under correlated observations with a given covariance function. One of them was proposed by Brimkulov et al. [Bri80], and has been adopted in Chapter 5 for MR designs. Unfortunately, the convergence of such an algorithm with the true optimum is not guaranteed. Müller and Pázman [Mü99] proposed an alternative algorithm based on a new interpretation of design measures [Paz98]. The restriction in this new interpretation is that replications are not allowed. We may be able to adopt that algorithm to MR designs and performing a comparison.

Another important issue is the measurement of how “optimal” a design is. This is measured by the efficiency but, as we saw, we need to know what the optimal design is in order to compute the efficiency; for this reason, we computed lower bounds. A lower bound of the efficiency has been proposed for covariance matrices which are symmetric, circular and nonsingular. (See Bischoff [Bis95]). From this arises a new issue. Could another bound be found for more general covariance matrix structure?

Finally other interesting issues might be:

- Application of the Bayesian theory for computing MR optimal designs.
- Elaboration of theories when new restrictions appear in the model and development of algorithms for solving them.
- Looking for conditions under which the sequential algorithm proposed in Chapter 4 converges.

Apéndice A

Apéndice

A.1. Cálculo de las derivadas locales de algunas funciones criterio

D–optimización

Sea

$$\Phi_D(\xi) \equiv \Phi_D[M(\xi)] = -\log \det\{M(\xi)\},$$

donde

$$M(\xi) = M(\xi_{2|1} \cdot \xi_1) = \sum_{i=1}^I \sum_{j=1}^{n_i} \xi_{2j|1i} [\xi_{1i} \eta(x_i, x_j) \eta^t(x_i, x_j)].$$

Como se vio en el capítulo 1

$$\begin{aligned} \frac{\partial \Phi_D}{\partial p_j} &= G_\Psi[M(\xi), \eta(x_j) \eta^t(x_j)] = -tr[M^{-1}(\xi) \eta(x_j) \eta^t(x_j)] \\ &= -\eta^t(x_j) M^{-1}(\xi) \eta(x_j). \end{aligned}$$

Escribiendo la matriz de información para los diseños marginalmente restringidos, como se hizo en el capítulo 3

$$\begin{aligned} M(\xi) &= \sum_i \sum_j \xi_{1i} \xi_{2j|1i} \eta(x_i, x_j) \eta^t(x_i, x_j) \\ &= \sum_i \sum_j \xi_{ij} \eta(x_i, x_j) \eta^t(x_i, x_j), \end{aligned}$$

se tiene que:

$$\frac{\partial \Phi_D}{\partial \xi_{2j|1i}} = \frac{\partial \Phi_D}{\partial \xi_{ij}} \frac{\partial \xi_{ij}}{\partial \xi_{2j|1i}} = \xi_{1i} \frac{\partial \Phi_D}{\partial \xi_{ij}} = -\xi_{1i} \eta^t(x_i, x_j) M^{-1}(\xi) \eta(x_i, x_j).$$

c -optimización

Sea

$$\Phi_c(\xi) \equiv \Phi_c[M(\xi)] = c^t M^{-1}(\xi) c,$$

en este caso

$$\begin{aligned} \frac{\partial \Phi_c}{\partial p_j} &= -tr[(M^{-1} c c^t M^{-1} \eta(x_j) \eta^t(x_j))] \\ &= -[\eta^t(x_j) M^{-1} c]^2, \end{aligned}$$

de donde resulta

$$\frac{\partial \Phi_c}{\partial \xi_{2j|1i}} = \frac{\partial \Phi_c}{\partial \xi_{ij}} \frac{\partial \xi_{ij}}{\partial \xi_{2j|1i}} = \xi_{1i} \frac{\partial \Phi_c}{\partial \xi_{ij}} = -\xi_{1i} [\eta^t(x_i, x_j) M^{-1} c]^2.$$

D_A -optimización

En este caso, la función criterio es

$$\Phi_{D_A}(\xi) = \Phi_{D_A}[M(\xi)] = \log \det\{A M^{-1}(\xi) A^t\},$$

donde A es una matriz $k \times m$ de rango $k < m$. De modo que

$$\begin{aligned}\frac{\partial \Phi_{DA}}{\partial p_j} &= -\text{tr}[(AM^{-1}A^t)^{-1}(-AM^{-1}\eta(x_j)\eta^t(x_j)M^{-1}A^t)] \\ &= -\eta^t(x_j)M^{-1}A^t(AM^{-1}A^t)^{-1}AM^{-1}\eta(x_j).\end{aligned}$$

Por tanto

$$\begin{aligned}\frac{\partial \Phi_{DA}}{\partial \xi_{2j|1i}} &= \frac{\partial \Phi_{DA}}{\partial \xi_{ij}} \frac{\partial \xi_{ij}}{\partial \xi_{2j|1i}} = \xi_{1i} \frac{\partial \Phi_{DA}}{\partial \xi_{ij}} \\ &= -\xi_{1i} \eta^t(x_1^{(i)}, x_2^{(j)}) M^{-1}(\xi) A^t [AM^{-1}(\xi) M^t]^{-1} AM^{-1}(\xi) \eta(x_1^{(i)}, x_2^{(j)})\end{aligned}$$

L_A -optimización

La función criterio es

$$\Phi_{LA}(\xi) = \Phi_{LA}[M(\xi)] = \text{tr}[AM^{-1}(\xi)A^t],$$

donde A es una matriz $k \times m$ de rango $k < m$, de modo que

$$\frac{\partial \Phi_{LA}}{\partial \xi_j} = -\eta^t(x_j)M^{-1}AA^tM^{-1}\eta(x_j).$$

Por tanto

$$\begin{aligned}\frac{\partial \Phi_{LA}}{\partial \xi_{2j|1i}} &= \frac{\partial \Phi_{LA}}{\partial \xi_{ij}} \frac{\partial \xi_{ij}}{\partial \xi_{2j|1i}} = \xi_{1i} \frac{\partial \Phi_{LA}}{\partial \xi_{ij}} \\ &= -\xi_{1i} \eta^t(x_1^{(i)}, x_2^{(j)}) M^{-1}(\xi) A^t AM^{-1}(\xi) \eta(x_1^{(i)}, x_2^{(j)})\end{aligned}$$

De manera análoga, calculamos $d_{3k|1i}$, teniendo en cuenta que:

$$\frac{\partial \Phi}{\partial \xi_{3|1}} = \xi_1 \sum_{x_2} \xi_{2|1,3} \frac{\partial \Phi}{\partial \xi_{1,2,3}}$$

A.2. Código fuente para el cálculo de diseños D –óptimos MR con observaciones correladas

En el capítulo 5 hemos visto que cuando las observaciones son correladas, el cálculo de diseños óptimos marginalmente restringidos puede resultar computacionalmente muy complejo. En ese capítulo realizamos la adaptación del algoritmo de Brimkulov y otros para el cálculo de diseños D –óptimos. En este apéndice escribimos los archivos (M-files) en MATLAB que han sido implementados para encontrar los diseños óptimos marginalmente restringidos de la sección 5.6. Nos basamos en los archivos proporcionados por [Uci04], que los podemos encontrar en:

<http://www.bepress.com/snbe/vol8/iss2/art13/>

Modificando parte del código proporcionado para el caso de diseños marginales:

- **principal.m** Esta es a la que se llama y desde la cuál se va a ir ejecutando el resto de funciones. Esta función no toma valores de entrada. Lo único que queda fijo en esta función es el diseño marginal y el valor de δ que determina la parada del algoritmo. Además creamos una red de puntos de valores para los diseños condicionales.

```
n=11
xi=diseñoinitial(n,0,30);
independiente=[1 1 1 1 1 1 1 1 1 1];
s1=[41 14 64 20 72 85 104 69 89 10 25];
grid=1:1:30;
opciones = optimset('Display', 'notify');
deltagrid=zeros(size(grid));
K=covstein(s1,xi);
Kinv=inv(K);
F=funcionbase(independiente,s1,xi);
M=F'*Kinv*F;
detM=det(M);
disp(detM) D=inv(M);
```

```

iter=0;

iptworst=1;
while 1
deltabest = -inf;
for irmvpt = 1: n
% Eliminacion del i-ésimo punto del diseño previo
[xirmv, sirmv, independientermv, Frmv, Krmv, Kinvmv, frmvstar]
= eliminar(irmvpt, xi, s1, independiente, F, K, Kinv);

for loop = 1: ptosred
deltagrid(loop) = -inclusion(grid(loop), xirmv, sirmv,
independientermv, Frmv, Kinvmv, D, frmvstar,s1,irmvpt);
end
[deltagridmax, imax] = max(deltagrid);
imax = max(min(imax, ngridpts - 1), 2);
[xnew, ndelta, exitflag] = fminbnd(@inclusion, grid(imax - 1),
grid(imax + 1), opciones, xirmv, sirmv, independientermv, Frmv,
Kinvmv, D, frmvstar,s1,irmvpt);
delta = -ndelta;
if delta > deltabest
deltabest = delta;
iptworst = irmvpt;
xaddbest = xnew;
end
end
if deltabest < epsi
disp('Se ha encontrado una solución óptima...')
disp([s1' xi'])
disp('det( $M^{-1}$ ):')
disp(1 / detM)
xi = sort(xi);
save xi.txt xi -ascii
save det.txt detM -ascii
break
end
[xirmv, sirmv, independientermv, Frmv, Krmv, Kinvmv, frmvstar]
= eliminar(iptworst, xi, s1, independiente, F, K, Kinv);

[xiadd, sladd, Fadd, Kadd, Kinvadd, Madd, detMadd, Dadd] =
adaptacion(xaddbest, xirmv,s1, sirmv, Frmv, Krmv, Kinvmv, M,

```



```

detM, D, frmvstar, iptworst);
xi = xiadd;
s1= sladd;
F = Fadd;
K = Kadd;
Kinv = Kinvadd;
M = Madd;
detM = detMadd;
D = Dadd;
iter = iter + 1;
disp(['iteration ', num2str(iter), ', det(M) = ', num2str(detM)])
end

```

- **diseño inicial.m** En esta función se establece de manera aleatoria un diseño condicional inicial. Toma como parámetros de entrada el número de puntos del diseño marginal y los extremos del intervalo del espacio de diseño de la variable controlable.

```

function xi = diseñoinicial(n, t0, tn)
xi = t0 + ceil(tn * rand(1, n));

```

- **covstein.m** Esta función es la encargada de construir la matriz de covarianzas, de acuerdo con la estructura, que en el ejemplo del viento irlandés encontramos en [Ste05b]. Como parámetros de entrada utiliza el diseño marginal y los puntos creados en la función del diseño inicial.

```

function K=covstein(s1,x) for i=1:length(x)
for j=1:length(s1)
if s1(i)==s1(j)
K(i,j) = (0,901 * abs(x(i) - x(j))1,544 + 1)-1;
else
K(i,j) = 0,968 * ((0,901 * abs(x(i) - x(j))1,544 + 1)-1 * exp(-(0,00134 *
abs(s1(i) - s1(j))) * ((0,901 * abs(x(i) - x(j))1,544 + 1)0,61/2)-1);
end

```

```
end
end
```

- **funcionbase.m** En esta función se determina la matrix del diseño, X .

```
function F = funcionbase(independiente,s1,x)
F = [independiente',s1' x'];
```

- **eliminar.m** Esta función elimina el punto que será intercambiado por otro que haga mayor el determinante de la matriz de información.

```
function [xnew, s1new, independientenew, Fnew, Knew, Kinvnew,
fxnstar] = eliminar(ipt, xi, s1, independiente, F, K, Kinv)
% Eliminación de un punto. Por simplicidad, el punto eliminado
se intercambia con el último punto del diseño
```

```
n = length(xi);
ind = 1: n - 1;
if ipt < n,
ind(ipt) = n;
end
xnew = xi(ind);
s1new = s1(ind);
independientenew = independiente(ind);
Fnew = F(ind, :);
Knew = K(ind, ind);
A = Kinv(ind, ind);
u = Kinv(ind, ipt);
r = Kinv(ipt, ipt);
Kinvnew = A - u * u' / r;
fxnstar = sqrt(r) * F(ipt, :)' + Fnew' * u / sqrt(r);
```

- **inclusion.m** Esta función trata de añadir un nuevo punto al diseño. Y entre todos ellos se elige el que maximiza $\Delta(i, t)$.

```
function negDelta = inclusion(x, xi, s1n, independiente, F, Kinv,
D, frmvstar, s1, irmvpt)
```

```

% Añadimos un nuevo punto al diseño
knx = covstein4(s1n,s1(irmvpt),xi,x)';

alpha = 1 - knx' * Kinv * knx;
if abs(alpha) < eps
fxstar = zeros(size(frmvstar));
else
fxstar = ([1 s1(irmvpt) x]' - F' * Kinv * knx) / sqrt(alpha);
end
dxstar = fxstar' * D * fxstar;
drmvstar = frmvstar' * D * frmvstar;
dxrmvstar = fxstar' * D * frmvstar;
negDelta = -(dxstar - drmvstar - drmvstar * dxstar + dxrmvstar
* dxrmvstar);

```

- **covstein4.m** Esta función nos proporciona la matriz de covarianzas añadiendo el nuevo punto.

```

function K=covstein4(s,l,x,y)
for i=1:length(s)
K(i) = 0,968 * ((0,901 * abs(x(i) - y)(1,544)+1)-1) *
exp(-(0,00134*abs(s(i)-l))*((0,901*abs(x(i)-y)(1,544)+1)(0,61/2))-1);
end

```

- **adaptacion.m** Una vez seleccionado el punto a añadir, mediante esta función construimos la “nueva” matriz de información, matriz de covarianzas, y la inversa de la matriz de información.

```

function [xnew, s1new, Fnew, Knew, Kinvnew, Mnew, detMnew, Dnew]
= adaptacion(x, xi, s1, s1n, F, K, Kinv, M, detM, D, frmvstar,
iptworst)

xnew = [xi, x];
s1new= [s1n, s1(iptworst)];
Fnew = [F; 1, s1(iptworst),x];
kxx = 1;

```

```

knx = covstein4(sln, s1(iptworst),xi, x)';
Knew = [K, knx; knx', kxx];
b = Kinv * knx; alpha = kxx - knx' * b;
fxstar = ([1 s1(iptworst) x]' - F' * b) / sqrt(alpha);
Kinvnew = zeros(size(Knew));
Kinvnew(1: end - 1, 1: end - 1) = Kinv + b * b' / alpha;
Kinvnew(1: end - 1, end) = -b / alpha;
Kinvnew(end, 1: end - 1) = Kinvnew(1: end - 1, end)';
Kinvnew(end, end) = 1 / alpha;
dxstar = fxstar' * D * fxstar;
drmvstar = frmvstar' * D * frmvstar;
dxrmvstar = fxstar' * D * frmvstar;
delta = dxstar - drmvstar - drmvstar * dxstar
+ dxrmvstar * dxrmvstar;
Mnew = M + fxstar * fxstar' - frmvstar * frmvstar';
detMnew = (1 + delta) * detM;
Dnew = D - D * (-(1 + dxstar) * frmvstar * frmvstar' + ...
dxrmvstar * (frmvstar * fxstar' + fxstar * frmvstar')) + ...
(1 - drmvstar) * fxstar * fxstar') * D / (1 + delta);

```

A.3. Diseño óptimo del problema de densificación en el caso correlado

Las Tablas A.1 – A.3 muestran el diseño óptimo conjunto para el problema de densificación cuando existe correlación entre las observaciones. (Véase sección 5.7.)

El valor del determinante de la matriz de información que se obtuvo fue $|M| = 5.6834 \times 10^{14}$.

s	t	s	t	s	t	s	t
94.89	19.80	95.29	8.50	95.33	3.32	95.37	7.30
95.07	7.41	95.29	19.03	95.33	4.95	95.39	17.16
95.11	2.80	95.29	12.17	95.33	12.42	95.39	19.13
95.11	0.39	95.29	0.54	95.33	18.71	95.39	8.19
95.15	16.92	95.29	5.84	95.33	8.37	95.39	9.72
95.15	6.09	95.29	9.03	95.33	0.39	95.39	7.06
95.15	12.33	95.29	9.81	95.33	8.58	95.39	6.25
95.17	4.62	95.29	1.19	95.33	10.83	95.39	2.27
95.20	0.00	95.31	4.06	95.33	17.07	95.39	9.56
95.21	11.50	95.31	18.64	95.33	16.36	95.39	6.72
95.22	15.82	95.31	10.06	95.33	12.14	95.39	2.94
95.22	16.59	95.31	19.80	95.33	1.07	95.39	9.27
95.23	11.25	95.31	20.00	95.33	5.43	95.40	8.00
95.23	11.92	95.31	1.84	95.34	12.14	95.41	15.39
95.23	15.24	95.31	12.41	95.34	1.16	95.41	14.76
95.23	0.20	95.31	6.29	95.35	17.40	95.41	14.94
95.24	15.09	95.32	5.46	95.35	12.22	95.41	16.76
95.24	3.23	95.32	16.24	95.35	7.59	95.41	6.06
95.25	9.32	95.32	12.33	95.35	13.96	95.41	12.90
95.25	14.19	95.32	6.58	95.36	15.84	95.41	12.82
95.25	17.20	95.32	16.30	95.36	3.98	95.41	3.42
95.25	0.26	95.32	19.12	95.36	13.63	95.42	4.62
95.28	10.74	95.32	16.92	95.36	3.02	95.42	17.87
95.28	0.88	95.32	10.50	95.36	13.20	95.43	17.83
95.28	13.60	95.32	3.79	95.36	8.69	95.43	18.44
95.28	1.78	95.32	9.03	95.36	18.18	95.43	16.26
95.28	17.97	95.33	9.97	95.37	4.02	95.43	0.31
95.28	11.94	95.33	9.48	95.37	0.58	95.43	3.48
95.28	13.40	95.33	7.13	95.37	12.31	95.43	12.52
95.28	11.91	95.33	8.18	95.37	5.44	95.43	12.20

Tabla: A.1 : Diseño D-óptimo MR para el ejemplo de densificación

s	t	s	t	s	t	s	t
95.43	14.03	95.53	9.09	95.64	18.17	95.74	12.42
95.43	8.88	95.53	16.66	95.64	5.95	95.74	19.25
95.43	1.42	95.53	7.05	95.64	9.70	95.75	11.36
95.44	4.05	95.53	4.92	95.65	16.38	95.75	15.78
95.44	3.87	95.54	3.53	95.66	5.79	95.75	8.80
95.44	18.00	95.54	8.90	95.66	14.35	95.75	14.55
95.45	11.75	95.54	12.25	95.66	16.89	95.75	5.10
95.45	16.77	95.54	1.57	95.66	19.99	95.75	19.89
95.45	3.76	95.55	0.32	95.67	17.86	95.75	2.74
95.46	19.00	95.56	16.88	95.67	1.18	95.76	14.54
95.46	8.21	95.56	6.80	95.68	0.20	95.76	0.11
95.46	16.64	95.57	1.68	95.68	17.81	95.77	10.45
95.47	8.11	95.58	9.02	95.69	3.80	95.77	11.66
95.47	13.44	95.58	2.80	95.70	8.24	95.77	6.82
95.47	13.64	95.59	7.86	95.70	7.21	95.77	13.30
95.47	16.43	95.59	11.05	95.71	16.77	95.77	18.40
95.47	12.60	95.60	14.05	95.71	1.86	95.77	16.46
95.48	8.80	95.60	19.66	95.71	9.57	95.77	5.05
95.49	7.51	95.61	6.84	95.72	10.68	95.77	4.52
95.49	12.80	95.61	1.00	95.72	11.10	95.78	13.51
95.49	3.50	95.61	3.82	95.72	8.45	95.78	13.75
95.49	16.43	95.62	3.82	95.72	9.81	95.78	9.15
95.49	3.97	95.62	3.32	95.72	6.70	95.78	12.50
95.50	8.89	95.62	11.42	95.72	13.48	95.79	7.41
95.50	12.08	95.63	0.33	95.73	8.90	95.79	16.06
95.50	7.35	95.63	2.42	95.73	7.84	95.80	3.07
95.51	18.34	95.63	10.98	95.73	13.30	95.80	6.28
95.52	1.15	95.63	19.23	95.73	5.24	95.80	9.22
95.52	2.29	95.64	0.71	95.74	17.31	95.81	3.11
95.53	12.17	95.64	12.63	95.74	9.01	95.81	17.12

Tabla: A.2: Diseño D-óptimo MR para el ejemplo de densificación (cont.)

s	t	s	t	s	t	s	t	s	t
95.83	11.34	95.93	14.32	96.03	4.28	96.09	2.39	96.26	17.40
95.83	11.95	95.93	4.79	96.03	4.25	96.09	18.68	96.27	6.19
95.83	8.64	95.93	6.92	96.03	17.88	96.09	15.53	96.27	12.55
95.84	0.20	95.93	11.07	96.03	4.18	96.10	14.75	96.27	14.65
95.84	12.69	95.95	13.00	96.03	1.01	96.10	15.21	96.28	14.63
95.85	11.36	95.96	2.73	96.03	12.87	96.11	6.10	96.28	10.27
95.85	13.85	95.96	10.60	96.03	16.78	96.11	19.20	96.28	1.44
95.85	13.98	95.96	17.41	96.04	9.22	96.11	2.68	96.31	11.31
95.85	14.80	95.96	13.98	96.04	19.80	96.11	14.70	96.32	18.10
95.86	10.93	95.96	14.01	96.04	5.36	96.11	8.40	96.32	17.47
95.86	15.88	95.98	17.46	96.05	3.98	96.12	5.43	96.35	20.00
95.86	0.99	95.98	3.88	96.05	8.27	96.12	8.66	96.37	9.79
95.87	15.07	95.98	12.04	96.05	14.07	96.12	0.24	96.37	5.07
95.87	18.63	95.99	7.36	96.06	8.84	96.12	0.00	96.38	19.25
95.87	14.02	96.00	19.65	96.06	17.60	96.12	3.46	96.40	16.64
95.88	7.07	96.01	3.98	96.06	8.47	96.12	10.31	96.40	1.68
95.88	13.86	96.01	7.60	96.07	15.88	96.13	16.10	96.44	4.64
95.88	7.31	96.01	0.30	96.07	8.77	96.14	12.92	96.44	17.81
95.90	1.18	96.01	17.49	96.07	18.67	96.15	5.29	96.46	11.89
95.90	9.97	96.01	14.53	96.07	5.97	96.15	15.01	96.48	8.44
95.90	13.67	96.01	11.83	96.07	8.31	96.16	11.38	96.57	6.92
95.90	10.97	96.02	7.43	96.07	6.40	96.16	19.23	96.60	13.89
95.91	15.90	96.03	19.59	96.07	12.58	96.17	13.23	96.60	15.78
95.91	1.30	96.03	6.68	96.08	7.94	96.18	15.67	96.66	3.21
								96.69	0.00
								96.73	20.00

Tabla: A.3: Diseño D-óptimo MR para el ejemplo de densificación
(cont.)

Bibliografía

- [Abr65] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions*. Dover, New York, 1965.
- [Abt98] M. Abt and W. J. Welch. Fisher information and maximum-likelihood estimation of covariance parameters in gaussian stochastic processes. *The Canadian Journal of Statistics*, 26(1):127–137, 1998.
- [Adl05] R. J. Adler and J. E. Taylor. *Random Fields and Geometry*. Birkhäuser, Boston (forthcoming), 2005.
- [Ard92] R. Ardanuy and J. López Fidalgo. Optimal design with constraint support. *Revista de Matemática e Estatística*, 10:193–205, 1992.
- [Atk92] A. C. Atkinson and A. N. Donev. *Optimum Experimental Designs*. New York: Oxford Science Publications, 1992.
- [Atw69] C. L. Atwood. Optimal and efficient designs of experiments. *The Annals of Mathematical Statistics*, 40(5):1570–1602, 1969.
- [Atw76] C. L. Atwood. Convergent design sequences, for sufficiently regular optimality criteria. *Annals of Statistics*, 4:1124–1138, 1976.
- [Atw80] C. L. Atwood. Convergent design sequences, for sufficiently regular optimality criteria, II: singular case. *Annals of Statistics*, 8:894–912, 1980.

- [Bal05] I. Ballesteros, J. López Fidalgo, R. Martín Martín, and B. Torsney. Construction of marginally and conditionally restricted designs in uranium production. *Proceedings of the 5th Annual ENBIS Conference, Newcastle*, 2005.
- [Bis95] W. Bischoff. Lower bounds for the efficiency of designs with respect to the D-criterion when the observations are correlated. *Statistics*, 27:27–44, 1995.
- [Bri80] U.N. Brimkulov, G.K. Krug, and V.L. Savanov. Numerical construction of exact experimental designs when the measurements are correlated (in russian). *Zavodskaya Laboratoria (Industrial Laboratory)*, 36:435–442, 1980.
- [Cal83] H. I. Calvete. *Diseño óptimo de experimentos: Diseños marginalmente restringidos*. Tesis doctoral, Facultad de Ciencias, Universidad de Zaragoza, 1983.
- [Che53] H. Chernoff. Locally optimal designs for estimating parameters. *The Annals of Mathematical Statistics*, 24:586–602, 1953.
- [Coo80] R. D. Cook and L. A. Thibodeau. Marginally restricted D-optimal designs. *Journal of the American Statistical Association*, 75:366–371, 1980.
- [Coo95] D. Cook and V. Fedorov. Constrained optimization of experimental design. *Statistics*, 26:129–178, 1995.
- [Cre93] N. Cressie. *Statistics for Spatial Data. Revised edition*. Wiley, New York., 1993.
- [Cur91] C. Currin, T.J. Mitchell, M.D. Morris, and K. Ylvisaker. Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *J. Amer. Statist. Ass.*, 86:953–963, 1991.
- [Dem77] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm.

- Journal of the Royal Statistical Society, Ser. B*, 39:1–38, 1977.
- [Dob83] A.J. Dobson. *An Introduction to Statistical Modelling*. Chapman and Hall, London, 1983.
- [Elf52] G. Elfving. Optimum allocation in linear regression theory. *The Annals of Mathematical Statistics*, 23:255–262, 1952.
- [Fed72] V. V. Fedorov. *Theory of optimal experiments*. New York: Academic Press, 1972.
- [Fed88] V. V. Fedorov and A. Atkinson. The optimum design of experiments in the presence of uncontrolled variability and prior information. Dodge Y., Fedorov V. V. y Wynn H. P. (eds.), *Optimal Designs and Analysis of Experiments*, Amsterdam: North-Holland, 1988.
- [Fed97] V. V. Fedorov and P. Hackl. *Model-Oriented Design of Experiments*. Lect. Notes in Statistics: Springer-Verlag New York, Inc., 1997.
- [Fed04] V. V. Fedorov and S. Leonov. Optimal designs for regression models with forced measurements at baseline. *MODA 7 – Advances in model-oriented design and analysis. Contrib. Statist., Physica, Heidelberg*, páginas 61–69, 2004.
- [Fel74] J. Fellman. On the allocation of linear observations. *Commentationes Physico-Matematicae, Helsinki*, 44:27–78, 1974.
- [For92] I. Ford, B. Torsney, and C. F. J. Wu. The use of canonical form in the construction of locally optimal designs for non-linear problems. *J. Roy. Statist. Soc. Ser.*, B54:569–583, 1992.
- [Gar02] S. Garcet Rodríguez. *Diseño óptimo de experimentos con variables no controlables*. Tesis doctoral, Facultad de Ciencias, Universidad de Salamanca, 2002.

- [Gne02] T. Gneiting. Nonseparable, stationary covariance functions for space–time data. *J. Amer. Statist. Ass.*, 97:590–600, 2002.
- [Gup85] R. Gupta and R. Richards. Testing optimality of experimental designs for a regression model with random variables. *Journal of Statistical Planning and Inference*, 11:75–80, 1985.
- [Hac95] P. Hackl. Optimal designs for experiments with potentially failing trials. *MODA 4: Advances in Model-Oriented Data Analysis*. Physica Verlag, Heidelberg., páginas 117–124, 1995.
- [Han94] M. S. Handcock and J. R. Wallis. An approach to statistical spatial-temporal modeling of meteorological fields. *J. Amer. Statist. Ass.*, 89:368–378, 1994.
- [Har74] D. A. Harville. Nearly optimal allocation of experimental units using observed covariate values. *Technometrics*, 16:589–599, 1974.
- [Her76] A. M. Herzberg and D. F. Andrews. Some considerations in the optimal design of experiments in non-optimal situations. *J. Roy. Statist. Soc. Ser. B*, 38:284–289, 1976.
- [Hoe58] P. G. Hoel. Efficiency problems in polynomial estimation. *The Annals of Mathematical Statistics*, 35:1553–1560, 1958.
- [Hua93] M. N. L. Huang and M. C. Hsu. Marginally restricted linear–optimal designs. *Journal of Statistical Planning and Inference*, 35:251–266, 1993.
- [Hua95] M. N. L. Huang and M. N. Chang. Marginally restricted constrained optimal designs. *Sankhyā: The Indian Journal of Statistics*, B57(1):128–145, 1995.
- [Imh01] L. Imhof, J. Lopez Fidalgo, and W. K. Wong. Efficiencies of optimal approximate designs for small samples. *Statistica Neerlandica*, 55:301–318, 2001.

- [Kar66] S. Karlin and W. Studden. Optimal experimental designs. *The Annals of Mathematical Statistics*, 37:783–810, 1966.
- [Kie59a] J. Kiefer. Optimum experimental designs. *Journal Royal Statistical Society*, Ser.B(21):272–319, 1959.
- [Kie59b] J. Kiefer and J. Wolfowitz. Optimum design in regression problems. *The Annals of Mathematical Statistics*, 30:271–294, 1959.
- [Kie60] J. Kiefer. Optimum experimental designs V, with applications to systematic and rotatable designs. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.*, 1:381–405, 1960.
- [Kie61] J. Kiefer. Optimum designs in regression problems, II. *The Annals of Mathematical Statistics*, 32:298–325, 1961.
- [Kie74] J. Kiefer. General equivalence theory for optimum designs (aproximate theory). *The Annals of Statistics*, 2:848–879, 1974.
- [Kie76] J. Kiefer and W. J. Studden. Optimal design for large degree polynomial regression. *The Annals of Statistics*, 4(6):1113–1123, 1976.
- [Lóp98] J. López Fidalgo and J. M. Rodríguez Díaz. Characteristic polynomial criteria in optimal experimental design. *MO-DA 5 Advanced in Model-Oriented Data Analysis and Experimental Design*, Physica-Verlag, páginas 31–38, 1998.
- [Lóp04a] J. López Fidalgo and S. Garcet Rodríguez. Optimal experimental designs when some independent variables are not subject to control. *J. American Statistical Association*, 99(468):1190–1199, 2004.
- [Lóp04b] J. López Fidalgo and R. Martín Martín. Diseños D-óptimos marginalmente restringidos en modelos de regresión lineal. *Actas del XXVIII Congreso Nacional de Estadística, Investigación Operativa*, Cádiz, 2004.

- [Lóp04c] J. López Fidalgo, J. M. Rodríguez Díaz, G. Sánchez, and M. T. Santos Martín. Optimal design for compartmental models with correlated observations. *Journal of Applied Statistics (in press)*, 2004.
- [Lóp04d] J. López Fidalgo and J.M. Rodríguez Díaz. Elfving method for computing c-optimal designs in more than two dimensions. *Metrika*, 59:235–244, 2004.
- [Lóp05] J. López Fidalgo and R. Martín Martín. Marginally restricted D-optimal designs for linear regression models. *Proceedings of the 25th European Meeting of Statisticians, Oslo*, 2005.
- [Mü99] W.G. Müller and A. Pázman. An algorithm for the computation of opitimum designs under a given covariance structure. *Comput. Statistics*, 14, 1999.
- [Mü03] W.G. Müller and A. Pazman. Measures for designs in experiments with correlated errors. *Biometrika*, 90(2), 2003.
- [Mü04] W.G. Müller and M. Stehlík. An example of D-optimal designs in the case of correlated errors. *COMPSTAT2004 Proceedings, Springer*, páginas 1519–1526, 2004.
- [Man00] S. Mandal and B. Torsney. Algorithms for the construction of optimizing distributions. *Communications in Statistics - Theory and Methods*, 29:1219–1231, 2000.
- [Man06] S. Mandal and B. Torsney. Construction of optimal designs using a clustering approach. *J. Stat. Planning and Inference*, 136(3):1120–1134, 2006.
- [Mat86] B. Matérn. *Spatial Variation*. 2nd edition. Springer, Berlin, 1986.
- [Nac89] C. J. Nachtsheim. On the design of experiments in the presence of fix covariates. *Journal of Statistical Planning and Inference*, 22:203–212, 1989.

- [Paz86] A. Pazman. *Foundations of optimum experimental design*. D. Reidel publishing company, Dordrecht, 1986.
- [Paz98] A. Pazman and W.G. Müller. A new interpretation of design measures. *MODA5, Physica Verlag*, páginas 239–246, 1998.
- [Paz04] A. Pazman. Correlated optimum design with parametrized covariance function: Justification of the use of the Fisher information matrix and of the method of virtual noise. *Report #5 of the Research Report Series of the Department of Statistics and Mathematics, Wirtschaftsuniversität Wien, Austria*, 2004.
- [Pro02] L. Pronzato. Information matrices with random regressors. application to experimental design. *Journal of Statistical Planning and Inference*, 108:189–200, 2002.
- [Puk93] F. Pukelsheim. *Optimal Design of Experiments*. John Wiley and Sons, New York, 1993.
- [Roc72] R. T. Rockafellar. *Convex analysis*. New Jersey: Princeton University Press., 1972.
- [Rod99] C. Rodríguez and I. M. Torreblanca. *Diseño óptimo de experimentos para modelos de regresión*. Universidad de Almería, Servicio de publicaciones, Almería, 1999.
- [Rod00] J. M. Rodríguez Díaz. *Criterios Característicos en Diseño Óptimo de Experimentos*. Tesis doctoral, Facultad de Ciencias, Universidad de Salamanca, 2000.
- [Sah00] M. Sahm and R. Schwabe. A note on optimal bounded designs. *Atkinson, A., Bogacka, B. y Zhigljavsky A. (eds.), Optimun Design 2000, Dordrecht: Kluwer*, páginas 131–140, 2000.
- [Sch96] R. Schwabe. *Optimun Design for Multi-Factor Models*. New York: Springer-Verlag, 1996.

- [Sch04] R. Schwabe and W.K. Wong. A relationship between efficiencies of Marginal Designs and the product design. *Metrika*, 45(3):253–257, 2004.
- [Sil78] S. D. Silvey, D. M. Titterington, and B. Torsney. An algorithm for optimal designs on a finite design space. *Communications in Statistics*, A, 14:1379–1389, 1978.
- [Sil73] S. D. Silvey and D. M. Titterington. A geometric approach to optimal design theory. *Biometrika*, 60(1):21–32, 1973.
- [Sil80] S. D. Silvey. *Optimal Design*. London: Chapman and Hall, 1980.
- [Sit93] R.R. Sitter and C.F.J. Wu. Optimal designs for binary response experiments: Fieller, D and A criteria. *Scand. J. Statist.*, 20:329–341, 1993.
- [Sit95] R.R. Sitter and B. Torsney. D-optimal designs for generalized lineal models. *Proceedings of MODA4-Model Oriented Data Analysis*, páginas 87–102, 1995.
- [Sme99] G.W. Smetana. Preoperative pulmonary evaluation. *Engineering Journal of Medicine*, 140:937–944, 1999.
- [Smi18] K. Smith. On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution observations. *Biometrika*, 12:1–85, 1918.
- [Ste04] M. Stehlík. Further aspects on an example of D-optimal designs in the case of correlated errors. *Research Report Series / Department of Statistics and Mathematics, Nr.1, University of Economics and Business Administration Viena*, 2004.
- [Ste05a] M. Stehlík, J. M. Rodríguez Díaz, W.G. Müller, and J. López Fidalgo. Optimal allocation of bioassays in the

- case of parametrized covariance functions: an application to lung's retention of radioactive particles. *Test (in press)*, 2005.
- [Ste05b] M. L. Stein. Space-time covariance functions. *J. Amer. Statist. Ass.*, 100:310–321, 2005.
- [Sze39] G. Szego. *Orthogonal Polynomials*. American Mathematical Society providence, Rhode Island, 1939.
- [Tit76] D. M. Titterington. Algorithms for computing D-optimal designs on a finite design space. *Conference on Information Sciencies and Systems, Dept. Elect. Eng., Johns Hopkins Univ. Baltimore, MD*:213–216, 1976.
- [Tor77] B. Torsney. Contribution to discussion of “maximum likelihood estimation via the EM algorithm” by Dempster et al. *J. Roy. Statist. Soc., B*, 39:26–27, 1977.
- [Tor83] B. Torsney. A moment inequality and monotonicity of an algorithm. *Proc. of the Internat. Symp. on SemiInfinite Prog. and Appl. (Edited by Kortanek, K.O. and Fiacco, A. V.) at the Univ. of Texas, Asutin. Lecture Notes in Econ. and Math. Sys.*, 215:249–260, 1983.
- [Tor88] B. Torsney. Computing optimizing distributions with applications in designs estimation and image processing. *Optimal Design and Analysis of Experiments. Editors: Y. Dodge, V.V. Fedorov and H.P. Wynn. North Holland*, páginas 361–370, 1988.
- [Tor92] B. Torsney and A.M. Alahmadi. Further development of algorithms for constructing optimizing distributions. *Model oriented data analysis, Proc. 2nd IIASA Workshop in St. Kyrik, Bulgaria, Physica-Verlag*, páginas 121–129, 1992.
- [Tor93] B. Torsney and A.K. Musrati. On the construction of optimal designs with applications to binary response and to

- weighted regression models. *Proceedings of the 3rd Model Oriented Data Analysis Workshop in Petrodvorets, Russia*, páginas 37–52, 1993.
- [Tor01] B. Torsney and S. Mandal. Construction of constrained optimal designs. *Atkinson, A. and Bogacka, B. and Zhigljavsky A. (eds.), Optimun Design 2000, Dordrecht: Kluwer*, páginas 141–152, 2001.
- [Tus81] A. V. Tuskanov. On test of optimality of experimental designs for a regression model with random variables. *Theory Prob. Applications*, 26:173–177, 1981.
- [Uci04] D. Uciński and A. C. Atkinson. Experimental design for time-dependent models with correlated observations. *Studies in Nonlinear Dynamics & Econometrics*, 8(2):13, 2004.
- [Var01] G. Varela, R. Cordovilla, M. F. Jiménez, and N.Ñova. Utility of standarized exercise oximetry to predict cardiopulmonary morbidity after lung resection. *European Journal of Cardio-thoracic Surgery*, 19:351–354, 2001.
- [Whi71] P. Whittle. *Optimization under Constraints*. Wiley, London, 1971.
- [Whi73a] L. V. White. An extension of the general equivalence theorem to non linear models. *Biometrika*, 60(2):345–348, 1973.
- [Whi73b] P. Whittle. Some general points in the theory of optimal experimental design. *Journal Royal Statistical Society, Ser. B*, 1:123–130, 1973.
- [Won95] W. K. Wong. On the equivalence of D- and G-optimal designs in heteroscedatisc models. *Statistics and Probability Letters*, 25:317–321, 1995.
- [Wu78] C. F. J. Wu and H. P. Wynn. The convergence of general step-length algorithms for regular optimum design criteria. *Annals of Statistics*, 6:1273–1285, 1978.

- [Wyn70] H. P. Wynn. The sequential generation of D-optimum experimental designs. *The Annals of Statistics*, 6(6):1273–1285, 1970.
- [Wyn72] H. P. Wynn. Results in the theory and construction of D-optimum experimental design. *Journal Royal Statistical Society, Ser. B.*, 34:133–147, 1972.
- [Wyn75] H. P. Wynn. Simple condicions for optimum design algorithms. *J.N. Srivastava, ed., A Survey of Statistical Design and Linear Models*, North-Holand Publishing Company, páginas 571–579, 1975.
- [Zha99] F. Zhang. *Matrix Theory: Basic Results and Techniques*. Springer-Verlag, New York, 1999.
- [Zha04] H. Zhang. Inconsistent estimation and asymptotitcaly equivalent interpolations in model-based geostatistics. *Journal of the American Statistical Association*, (99):250–261, 2004.
- [Zhu04] Z. Zhu and M. L. Stein. Spatial sampling design for parameter estimation of the covariance function. (*In press*), 2004.